GEOVISUALIZATION OF KNOWLEDGE DIFFUSION: VISUALIZATION OF BIBLIOGRAPHIC DATA 1995-2009

A THESIS PRESENTED TO
THE DEPARTMENT OF GEOLOGY AND GEOGRAPHY
IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE

By
David E. Hubbard

NORTHWEST MISSOURI STATE UNIVERSITY
MARYVILLE, MISSOURI
March, 2012

Geovisualization of Knowledge Diffusion:

Visualization of Bibliographic Data 1995-2009

David E. Hubbard

Northwest Missouri State University

THESIS APPROVED

Thesis Advisor, Dr. Patricia Drews Date

Dr. Yanfen Le Date

Dr. Yi-Hwa Wu Date

Dean of Graduate School Date

Geovisualization of Knowledge Diffusion:

Visualization of Bibliographic Data 1995-2009

Abstract

Bibliometrics are an important research area within information and library science, which provides valuable insights about relationships between authors, publications, and knowledge domains. This study examined the geographic aspects of literature involving the visualization of bibliographic data published by authors residing in the contiguous United States. It determined where visualization of bibliometric research occurred and explored the spatial relationships among its contributors via institutional affiliation. The study involved five aspects: (1) cited publications, (2) citing publications, (3) cited-citing publication networks, (4) co-author networks and distances, and (5) hypothesis testing of average co-author distances over time.

Using 102 publications identified from Thomson Reuters' Web of Science in the field of visualization of bibliographic data, it demonstrated that spatial aspects of bibliographic data can be represented in ArcGIS as both points (institutions) and networks (cited-citing pairs). The study examined clustering of the bibliographic data based institutional affiliation (i.e., ZIP code) using a nearest neighbor analysis. A Visual Basic for Applications (VBA) script was used to create polylines for cited-citing publication and co-author networks. The networks were mapped using small multiples and animation. Average co-author distances were calculated for the co-author networks and temporal changes were explored formally using a nonparametric hypothesis test. The average nearest neighbor analysis found that both cited and citing publications involving visualization of bibliographic data were clustered. Visual inspection of the thematic maps showed clustering of both cited and citing maps concentrated in the following cities: Philadelphia, PA, Bloomington, IN, Sandia, NM, Stillwater, OK, and Tucson, AZ. Despite a statistically significant increase in the number co-authored publications on visualization of bibliographic data, there was no change in the average co-author distances from 2001-2009.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my wife for all her love and encouragement throughout this program and every day. I wish to express my sincere gratitude to Dr. Patricia Drews for her advice and guidance as my thesis advisor. Her willingness to explore such a novel application of GIS was greatly appreciated. I also appreciate the efforts and suggestions of Drs. Yanfen Le and Yi-Hwa Wu who served on my thesis committee. Last but not least, to my family for their support and understanding during my GIS program.

Chapter 1

Introduction

Bibliometrics are an important research area within information and library science, which provides valuable insights about relationships between authors, publications, and knowledge domains. This study examines the geographic aspects of literature involving the visualization of bibliographic data published by authors residing in the contiguous United States. It demonstrates how ArcGIS can be utilized to represent spatial relationships among researchers in the field of visualization of bibliographic data, as well as perform spatial analyses to identify patterns and trends within that field.

1.1. Background

Scientific literature has long been viewed as a network, linking one paper to another through references and subsequent citations in the literature (Adair 1955, Garfield 1955). These networks became more obvious with the development of citation indexes in the 1960s that systemically indexed the journal literature and documented "who cited who" for the literature indexed. An entire field of study developed involving the use of citation indexes called citation analysis or more broadly bibliometrics.

Efforts to visualize citation data were contemporaneous with the development of citation indexes (Allen 1960), but according to Borner (2010) science mapping based on knowledge domains began in the 1930s-1940s. The graphic representations of citation data are called citation maps and have evolved as methods and software advanced. These maps show the relationship between authors, co-authors, and/or knowledge domains.

The purpose of these maps is to understand the relationships and trends among researchers, as well as the knowledge domains. The latter plays an important role in discovering new research trends and frontiers. Geographic aspects of bibliographic data have also been used to identify institutions with high research productivity, as well as study various types of collaborations (e.g., co-authorship, R&D spillovers, etc.). Unlike cartographic maps, citation maps typically have no geographic component and are mapped in relative space. Since researchers are usually affiliated with an institution, bibliographic data is inherently geographic.

Only a few studies involving citation maps explore the spatial relationships among researchers, institutions, and/or knowledge domains using spatial analyses (Hoekman *et al*. 2010, Maggioni and Uberti 2009, Waltman *et al*. 2011). An understanding of the geographic aspects of scholarship may be overdue considering how the Internet has revolutionized and facilitated communications over distance. Some have even referred to this phenomenon as the "death of distance" suggesting that this change essentially eliminates distance and impediments it brings to society (Cairncross 2001). This would suggest that collaboration and co-authorship might be changing since distance no longer matters and would not be an impediment to collaboration. In a study of scientific teams in the United States from 1981-1999, Adams *et al.* (2005) found increasing average co-author distances among major scientific disciplines. Waltman *et al*. (2011) found similar trends in the sciences and social sciences among a number of countries. However, others have found the "death of distance" to be premature (Hoekman *et al*. 2010, Maggioni and Uberti 2009), finding collaboration to be localized. The First

Law of Geography (Tobler 1970) would also suggest that co-authorship might remain localized despite a telecommunications revolution.

The geographic aspect of bibliographic data is beginning to be explored through geovisualization, which is an emerging field embracing a wide variety of novel mapping techniques. These mapping techniques involve dynamic and interactive aspects that allow one to see data in new ways (Dykes 2008). The field is closely allied with information visualization, but unlike information visualization, geovisualization always possesses a spatial component. Geovisualization has been applied to many standard cartographic applications (e.g., thematic maps), as well as more experimental or novel approaches that expand our ideas of mapping and communicating spatial information (e.g., mapping cyberspace). Geovisualization is not without its detractors, and some question their effectiveness compared to static maps, especially map animation.

1.2. Justification and Rationale

Citation analysis is a major research area within information and library science with over 1,500 published studies since the 1960s. This research expands upon that work by using GIS to explore the spatial aspects of bibliographic data, yielding a novel contribution to a large body of literature. It explores new ways to visualize bibliographic data not available using relative space. To date, there are no known studies that have mapped spatial aspects of bibliographic data using spatial analyses tools in ArcGIS. This research would serve as a template for those interested in creating citation maps in geographic space, as well as performing spatial analyses.

1.3. Research Objectives

The purpose of this study is to use ArcGIS for the geovisualization and spatial analysis of bibliographic data. There are two primary objectives of this study. The first is to determine if it is possible to represent the spatial aspects of bibliographic data in ArcGIS as points and polylines. The second objective is to determine if geovisualization and spatial analysis provide any additional insights compared to citation maps in relative space. One new insight compared to relative space is the examination of the "death of distance" among co-authors, which will be formally tested and compared to the few studies that have examined "death of distance" among co-authors. While any knowledge domain could be used to illustrate the proposed approach, this study examines journal and proceedings literature involving the visualization of bibliographic data from 1995 to 2009.

Chapter 2

Literature Review

2.1. Geovisualization

Advances in computer hardware, software, and an increased understanding of human cognition have contributed to the development of information visualization during the last two decades of the 20th century. More recently, geovisualization has emerged as an interdisciplinary field that draws upon many disciplines for the visual exploration, analysis, synthesis and presentation of data with spatial aspects (MacEachren and Kraak 2001). Geovisualization and information visualization share many attributes. Both disciplines employ an interdisciplinary approach and are interested in representing information in new ways, often resulting in tools and outputs that are interactive, multimodal, and experimental. While these two disciplines share a number of attributes, geovisualization is inherently spatial which presents a number of unique challenges and opportunities.

Map animation is one type of geovisualization. The benefits and the effectiveness of map animation have been challenged by Tversky *et al*. (2002), but this may depend on how one defines effectiveness and the purpose of the animation (Slocum *et al*. 2005, Dodge *et al*. 2008). In addition to the standard visual variables (e.g., spacing, size, orientation, shape, hue, location, etc.) first outlined by Bertin (1983), animated maps may also utilize display date, order, rate of change, duration, frequency, and synchronization (MacEachren 2004). In reviewing dynamic visual variables of maps, MacEachren (2004, p. 280) noted that "… on a dynamic map things that change attract more attention than things that do not….[and] time gives the map designer a powerful new graphic tool." He also stated that "[t]he simplest application of controlling period duration is in the binary cycling of on-off used in blinking .…[and] the value and/or hue of a map mark is changed back and forth between two states to draw attention to a place" (MacEachren 2004, p. 282) . The selection of duration, amount of time a frame (e.g., a map) is displayed, depends on the task and purpose of the animated map, but Griffin *et al*. (2006) found participants had the highest success rate identifying cluster movements using 1.5 seconds per frame in a cluster map study.

2.2. GIS and Geovisualization of Bibliographic Data

Most of the studies involving geovisualization of bibliographic data have occurred over the last ten years. This is to be distinguished from what Leydesdorff and Persson (2010) refer to as cognitive mapping, which occurs in relative rather than geographic space. Few of these studies have examined the spatial aspects of bibliographic data using a commercial GIS. When the spatial aspects are explored, it is often limited to determining the research productivity of institutions or nations. Many of these studies primarily use GIS to create graphics and spatial analyses are often nonexistent.

One of the earliest and more novel applications of using GIS to map bibliographic data involved cognitive (i.e., relative space) rather than geographic mapping. Old (2001) utilized GIS to create contour and 3-D maps of co-citation counts in ESRI's ArcGIS. A multidimensional scaling method was utilized to map 75 canonical information science authors. His objective was to illustrate how spatial methods can be used to map non-spatial data in novel ways. While Old stated that he wanted "to show how spatial analysis can contribute to the researcher’s set of analysis tools” (Old 2001, p. 564), his conclusion outlined some limitations and mentioned further analysis was needed. One of the limitations he mentioned was dealing with time series, which this thesis explores. More importantly, Old (2001) was the first to employ ArcGIS in the study of bibliographic data and paved the way for others, including this thesis.

Skupin (2002, 2004) also used GIS to map bibliographic data in relative space. More specifically, he created knowledge domain visualizations using ArcGIS from abstracts of the 1999 Association of American Geographers annual meeting. He accomplished this using an artificial neural network to create self-organizing maps (or Kohonen maps) and ultimately Thiessen polygons based on keywords identified. Various clustering methods were used to aggregate keywords into broader topics to obtain large format domain knowledge maps. In all three studies, Old (2001) and Skupin (2002, 2004) applied cartographic methods and GIS to map bibliographic data in relative space.

The earliest studies to map bibliographic data in geographic space utilized hand drawn/traced maps (e.g., Matthiessen and Schwartz 1999) and networks showing a “natural geographic order” (e.g., Glanzel 2001). However, Batty (2003) conducted one of the first studies that examined the spatial aspects of bibliographic data using GIS. The study mapped the most highly cited papers from Institute for Scientific Information’s HighlyCited database based on institutional affiliation of the first (or primary) author. Batty found that 40% of the most highly cited authors worked at ten institutions, nine of which were in the United States. The results were then presented as thematic maps with proportional symbols, which is a typical approach when GIS is employed to present bibliographic data (i.e., primarily limited to presentation of the results).

In a similar manner, Borner *et al*. (2006) used a set of papers published in the Proceedings of the National Academy of Sciences (PNAS) from 1982-2001 to determine knowledge diffusion among U.S. institutions based on citations to and from PNAS papers. The authors determined that the top producers (i.e., institutions receiving most of the citations for their PNAS papers) were also the top consumers (i.e., citing most of the PNAS papers). The numbers of cited versus citing PNAS papers were mapped using ArcGIS and presented as thematic maps using proportional symbols. The authors also presented the flow of scholarly knowledge between top producers and consumers using an open source Perl application, Chizu (Meiss 2005), which generated flow maps of scholarship on a contiguous map of the United States.

2.3. In-House/Open Source Solutions versus Commercial GIS Software

When GIS is used to present spatial aspects of citation data, researchers often use open source or an in-house solution rather than commercial GIS software. The reason for this is unclear. It may stem from limitations of commercial GIS software for this application or the desire to develop something novel that caters to mapping citation data. The latter provides an opportunity to develop and market software, especially if the software can easily identify research trends in financially lucrative fields.

The work of Carvalho and Batty (2006) is one example of an open source approach to mapping bibliographic data. In their study, the authors created a map using a C program developed by Gastner and Newman (2004) to show the productivity of research centers in the conterminous United States using bibliographic data (i.e., metadata from over 716,000 computer science articles) from the Citeseer database.

Using a diffusion transform, cartograms were created factoring in R&D expenditure by state, population by state, and county. The cartograms aided in illustrating the uneven distribution of research productivity.

Mothe *et al*. (2006) used Tetralogie**,** a Unix-based software solution developed by the Institut de Recherche en Informatique de Toulouse, which combines data mining, GIS, and visualization into a single platform. The platform was developed to identify trends and emerging fields for competitive intelligence purposes. Data was obtained from a bibliographic database though any textual dataset with a spatial component could be mined. Tetralogie provides tools to create choropleth maps at the country level, as well as explore co-author networks, histograms, and spreadsheets independent of the choropleth maps. Interactivity is limited to temporal exploration of the choropleth maps and co-author networks.

The advent of Google Maps and Google Earth offered a more accessible GIS for a variety of applications, including researchers interested in mapping bibliographic data. Chaomei Chen and his colleagues have used Google Earth to map several co-author networks (Chen 2007, Chen *et al*. 2007, Chen *et al*. 2008) involving research on Avian Flu, data and knowledge engineering, and terrorism. In a similar manner, Leydesdorff and Persson (2010) employed both Google Maps and Google Earth to map co-author networks for articles in a core list of information science journals. In addition to networks, Google Maps has been used to represent authorship associated with the number of Medline publications (LaRowe *et al*. 2009) and abstracts presented at the annual meetings of the Society for Neuroscience (Lin *et al*. 2008) using proportional symbols.

Considering the number of papers using Google Earth and Google Maps to map bibliographic data, these tools appear to be the preferred GIS for such applications.

Geovisualization of bibliographic data has emerged as a distinct research area over the last ten years, yet the use of commercial GIS software (e.g., ArcGIS) has been limited to data presentation. Current approaches provide users with limited, if not nonexistent, ability to perform spatial analyses. There is an opportunity to demonstrate that commercial GIS software can contribute more to the geovisualization than previous efforts, including spatial analyses.

Chapter 3

Conceptual Framework and Methodology

3.1. Overview and Study Area

The objective of this study is to explore the use of GIS for the visualization and spatial analysis of bibliographic data. The overall methodology and main outputs are outlined in Figure 1. Like most bibliometric studies, this model assigns a single

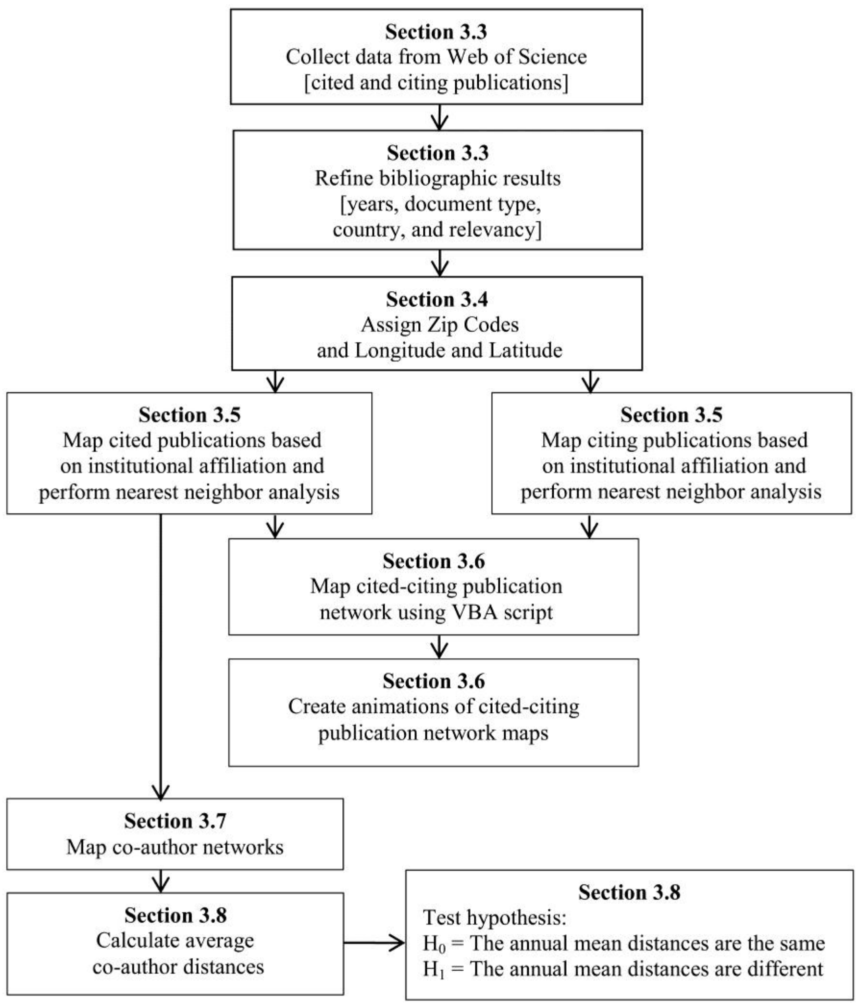

Figure 1. Overview of methodology and outputs. (Numbers refer to chapter sections.)

institutional affiliation based on the concept of first author, more specifically the ZIP code of the institution. While the study area could be global, the scope of this study was limited to the contiguous United States. This is an extent used in many previous studies (e.g., Borner *et al*. 2006, Carvalho and Batty 2006, LaRowe 2009).

3.2. Data Sources

The United States base map used in this study was obtained from the ESRI Data & Maps Series provided with ArcGIS 9.3 (ESRI 2008). The map originally possessed a GCS WGS 1984 Datum, but was transformed to NAD 1983 and projected as USA Contiguous Equidistant Conic in ArcGIS.

The bibliographic data used in this study was obtained from Thomson Reuters' Web of Science with ISI Proceedings (Thomson Reuters 2010). The initial data set was comprised of a citation (i.e., author, title, source, volume/issue, and pages), abstract, affiliation, and address.

Latitude and longitude were obtained for ZIP codes using ZIPList5 Geocode (http://www.zipinfo.com/search/zipcode.htm), which is a free service providing coordinates for the 2010 ZIP code centroids. The latitude and longitude are given in degrees to four decimal places and provide a "…general level of accuracy of under 100 feet" according to the provider CD Light (2010).

A Visual Basic for Applications (VBA) script was obtained from the ESRI Developer Network (ESRI 2004) to create polyline shapefiles for cited-citing and co-author networks using two pairs of latitudes/longitudes. The VBA script is in the public domain and shown in Appendix A.

3.3. Acquisition of Bibliographic Data

Thomson Reuters' Web of Science, a multidisciplinary citation database, indexes journal articles, proceeding papers, and other types of publications selectively. Within the Web of Science database, each publication indexed (i.e., cited publication) is linked to subsequent publications that reference or cite the publication (i.e., citing publication). In this thesis, cited publications refer to the initial Web of Science search results, and the citing publications refer to the publications that cite those publications.

Journal articles, proceeding papers, and review articles involving the visualization of bibliographic data were obtained from Thomson Reuters' Web of Science, a multidisciplinary citation database, using the following keyword search strategy: (visualiz* or mapping) and (citation or bibliomet* or "domain knowledge" or "knowledge domain*" or "subject domain*" or discipline). The results were refined within Web of Science by Publication Year (1995-2009), Document Type (Articles, Proceedings, and Reviews), and Countries/Territories (USA). The metadata (i.e., citation, abstract, affiliation, and address) of the resulting publications were exported to EndNote citation management software for management.

The full-text of exported publications was obtained and examined for relevance. The criterion for inclusion as a cited publication in this study was an actual visualization of bibliographic data within the publication itself (e.g., co-author citation map, subject cluster maps, mapping of bibliographic data, etc.) versus just standard tables and graphs. Since the study area was limited to the contiguous United States and co-author distance was determined, cited publications were refined further to include only those with authors and co-authors residing in the United States.

Publications citing the cited publications were downloaded from Web of Science as a tab-delimited text files. The text files were then imported into Excel for management. The citing publications were limited to those having at least the first author residing in United States. A unique publication code was then added to connect the citing publications to the original cited publications in the spreadsheet. The publication code was created using a combination of the first 4 letters of the author's surname, first 4 letters of the publication title (omitting a, an, and the), and the volume (or year if no volume number is available) of the cited publication and assigned to their respective citing publications.

3.4. ZIP Code to Latitude/Longitude

The ZIP code associated with the first author's institution was used throughout the study for the geographic location. In instances where Web of Science provided a ZIP code for the first author, it was utilized as the institution's address. If not present, ZIP codes were obtained from the actual publications. In a few instances where neither Web of Science nor the publication provided a ZIP code for the first author, the institution's webpage was searched on the Internet and ZIP code located.

The latitude and longitude of each institution was determined from the ZIP code based on the 2010 centroid. This was accomplished by individually looking up ZIP codes in ZIPList5 Geocode (CD Light 2010). The coordinates were determined to four decimal places and then added to their respective Excel spreadsheets.

3.5. Cited and Citing Publications

A United States base map from ESRI was added to ArcMap, which was transformed from GCS WGS 1984 to GCS NAD 1983 and then projected as USA Contiguous Equidistant Conic in ArcGIS. The United States base map was downloaded from ESRI® Data & Maps Series provided with ArcGIS 9.3 (ESRI 2008). ArcToolbox was used to transform the geographic coordinate system from WGS 1984 to NAD 1983 datum and to project the map to the USA Contiguous Equidistant Conic projection. The attribute table was edited to remove unnecessary fields (e.g., demographic data) and features (i.e., Alaska and Hawaii) were removed.

The Excel spreadsheets containing data for the cited and citing publications were edited to contain the fields shown in Table 1 and Table 2, respectively. Point shapefiles indicating institutions (i.e., ZIP code centroid associated with the first author) were created from Excel spreadsheets by first saving each file as a .csv file and then importing into ArcGIS using the "Add XY Data" tool. During the importing step, the geographic

Table 1. Spreadsheet fields for cited publications.

Field	Description
First_Au	First author listed on the cited publication
Inst	Institutional affiliation of the first author listed on the cited publication
Zip_Code	ZIP code of the institution
State	U.S. state where the institution is located
Latitude	Latitude in decimal degrees for the ZIP code centroid associated with the institution
Longitude	Longitude in decimal degrees for the ZIP code centroid associated with the institution
Cited_Pub	Code assigned to the cited publication

Table 2. Spreadsheet fields for citing publications.

Field	Description
First_Au	First author listed on the citing publication
Inst	Institutional affiliation of the first author listed on the cited publication
Zip_Code	ZIP code of the institution
State	U.S. state where the institution is located
Latitude	Latitude in decimal degrees for the ZIP code centroid associated with the institution
Longitude	Longitude in decimal degrees for the ZIP code centroid associated with the institution
Citing_Pub	Code assigned to the citing publication
Cited_Pub	Code assigned to the cited publication

coordinate system was defined as NAD 1983. The .csv files were converted to shapefiles in ArcMap and projected to the USA Contiguous Equidistant Conic projection. The symbology of the cited/citing icons was then edited to provide sufficient contrast for the points. A nearest neighbor analysis using the Euclidean Distance method in ArcToolbox was then conducted on both the cited and citing point shapefiles to examine clustering.

Using data contained in the two .csv files for cited and citing publications described above, two additional .csv spreadsheets were prepared that contain cumulative counts for each institution. These .csv files were then added to ArcGIS and converted to shapefiles in a process identical to the two point shapefiles above. Cumulative counts for cited and citing publications were presented as Bar/Column symbols on separate shapefiles.

3.6. Cited-Citing Publication Networks

Cited-citing publication networks were created by connecting each cited publication to its citing publications using polylines. The cited-citing publication networks were developed to present a geovisualization of the relationship between cited publications and their citing publications; it illustrates knowledge diffusion. This was accomplished by creating a spreadsheet for each of the publications that were cited at least once. Of those publications, only citing articles having a first author affiliated with a United States institution were retained. The institutional affiliation of the cited publication and their respective citing publications were then moved to a combined spreadsheet. Pairs of coordinates were then added for each pair of cited-citing publications, as well as year cited, ZIP codes, states, and institutional affiliations. Table 3 shows the fields used in the spreadsheet. The spreadsheet was resorted based on the year of the citing publications. Polyline shapefiles were created by transferring the pairs of coordinates (i.e., two x,y coordinate pair for each of the cited and citing publications) for each of the fifteen years from 1995 to 2009. The .csv spreadsheets were then saved as .dbf files. From ArcCatalog, a VBA script was run on the fifteen .dbf files to create polyline files of the cited-citing publications for each year.

Once the shapefiles were created in ArcCatalog, the geographic coordinate system was defined and the shapefiles were added to ArcMap. The shapefiles were then projected as USA Contiguous Equidistant Conic. Additional attributes were added, which included all the fields shown in Table 3 except for the latitude (CitingLat, CitedLat), longitude (CitingLong, CitedLong), and format (Format). A United States base map from

ESRI was added to ArcMap, which was transformed to a NAD 1983 and projected as USA Contiguous Equidistant Conic in ArcGIS.

Table 3. Spreadsheet fields for cited and citing publications.

Field	Description
Citing_Pub	Code assigned to the publication
CitingInst	Institutional affiliation of the first author listed on the citing publication
Citing_ZC	ZIP code of the citing institution
CitingLat	Latitude in decimal degrees for the ZIP code centroid associated with the institution
CitingLong	Longitude in decimal degrees for the ZIP code centroid associated with the institution
Citing_Au	First author listed on the citing publication
Cited_Pub	Code assigned to the cited publication
CitedInst	Institutional affiliation of the first author listed on the cited publication
Cited_ZC	ZIP code of the institution
CitedLat	Latitude in decimal degrees for the ZIP code centroid associated with the institution
CitedLong	Longitude in decimal degrees for the ZIP code centroid associated with the institution
Cited_Au	First author listed on the cited publication
Format	Document format of the citing publication (e.g., journal article, proceedings paper, etc.)

Three approaches for assigning hues to the polylines were explored for this map and others throughout this study. These were advance–retreat, single hue, and the unique values method. The advance-retreat approach takes advantage of the fact that longer wavelength hues (e.g., red) are perceived to be closer than shorter wavelength hues (e.g., blue) by the human eye. So when red and blue are used simultaneously, red appears in the foreground and blue in the background. The single hue approach employs a single hue (blue) for all polylines, and the unique values method assigns a unique hue to each polyline.

For the cited-citing publication network, a map was created with all fifteen cited-citing layers (one for each year) and the contiguous United States map layer. This allowed each year to be represented in a different hue. A unique values method was employed by assigning the following values to the red, green, and blue channels for each year: 0,0,255; 0,255,0; 255,0,0; 0,255,255; 255,0,255; 255,255,0; 0,0,128; 0,128,0; 128,0,0; 0,128,128; 128,0,128; 128,128,0; 0,0,64; 0,64,0; 64,0,0. The cited-citing publications were also presented as a series of small multiple maps, which is a method advocated by Bertin (1981) and Tufte (1990) for revealing pattern and change in a time series. Each of the small multiple maps represented a single year from 1995 to 2009. Map elements were added to the map layouts and exported as a jpegs.

Another set of small multiples was created using a single hue (blue) for all years, where each small multiple includes layers from the previous year (i.e., cumulative). Map elements were added to the map layout, including date display and exported as a jpegs. The results were then presented as both a series of small multiples and a 25-second animation to visualize the evolution of the cited-citing network over the 15-year period. The animation was created from those 15 jpegs using Microsoft’s Movie Maker and saved as a .wmv file. Another cumulative animation was created using the advance-retreat approach, where the current year’s cited-citing publication networks were assigned red and previous year’s blue. The two .wmv files are included in the Supporting Information.

3.7. Co-author Networks

Co-author networks were created by selecting papers with two or more authors from the original set of cited publications. This was accomplished by creating a separate layer for each publication with two or more authors. Geographic coordinates of the co-authors were determined using the same method as the first author approach described earlier. After obtaining coordinates for first authors and their co-authors, a spreadsheet was created with those coordinates for each first author-coauthor x,y coordinate pair using the same method as the cited-citing network. Fields were then added to the attribute table for each shapefile (Table 4). The distance between co-authors was calculated using the Calculate Geometry tool to obtain the distances. A single hue (blue) was used for a cumulative co-author network and the unique value method for the maps using small multiples for each year from 1995 to 2009. Appropriate map elements were then added to the maps and exported as jpegs

Table 4. Co-author network fields.

Field	Description
Cited_Pub	Code assigned to the cited publication
CitedInst	Institutional affiliation of the first author listed on the cited publication
Cited_Au	First author listed on the cited publication
CoAuth	Co-author listed on the cited publication
CoAuthInst	Institutional affiliation of the co-author listed on the cited publication

3.8. Co-author Distances

An average distance for each co-authored publication was calculated by summing the geographic distance from the first author to each co-author and then dividing by (n-1) co-authors. In cases where the first author and co-author had the same institutional affiliation or ZIP code, distance between those co-authors is zero. The number of co-authored papers was plotted for the fifteen-year period.

An annual mean distance was calculated from the average distances for each of the fifteen years. Normality of data within each of the fifteen years was then tested using Shapiro-Wilk. Because this test showed that the data were not normally distributed, the following hypothesis was tested using the Kruskal-Wallis test, which is the nonparametric version of Analysis of Variance, to determine if the distance between co-authors has changed over the fifteen-year period. The hypothesis tested is stated as:

H_0 = The annual mean distances are the same.

H_1 = The annual mean distances are different.

Where H_0 = Null Hypothesis and H_1 = Alternate Hypothesis

If the annual mean distances are different, Pearson's correlation analysis will be performed to explore the trend at the 0.05% significance level. With the exception of using Excel to plot of number of co-authored publications over time, all statistical analyses (i.e., Shapiro-Wilk, and Kruskal-Wallis) were performed using SAS 9.1.3.

Chapter 4

Analysis Results and Discussion

4.1. Bibliographic Data

The keyword search in Web of Science resulted in 1,243 publications, which were refined within Web of Science by Publication Year (1995-2009), Document Type (Article, Proceedings Paper, and Review), and Countries/Territories (USA) yielding 392 publications. After examining each publication for relevance (i.e., includes visualization of bibliographic data) and ensuring all authors and co-authors reside in the United States, there were 102 publications remaining. The final set of 102 cited publications is listed in the Appendix B.

Of the 102 cited publications, 60 were cited one or more times. The 60 publications were cumulatively cited by 1,357 publications; however, only 591 had a first author affiliated with a U.S. institution. Unlike the cited publications where all authors and co-authors were restricted to U.S. affiliations due to subsequent co-author analyses, citing publications were only required to have the first author affiliated with a U.S. institution. Web of Science primarily indexes journal articles and proceeding papers, but the 591 citing publications could potentially be from any source and their document types are summarized in Table 5.

Table 5. Format of citing publications.

Format	Count
Journal Article	405
Proceedings Paper	135
Book Chapter	50
Magazine	1

4.2. Maps of Cited and Citing Publications

The geographic locations of both cited and citing were mapped as points based on ZIP code. The resulting maps are presented in Figure 2 and Figure 4. The fields associated with the cited and citing publications are presented in Figure 3 and Figure 5, respectively. Publication codes (Cited_Pub) for the 102 cited publications are listed in Appendix B. The citing publication fields include a Cited_Pub field to connect the citing publication back to the original cited publication.

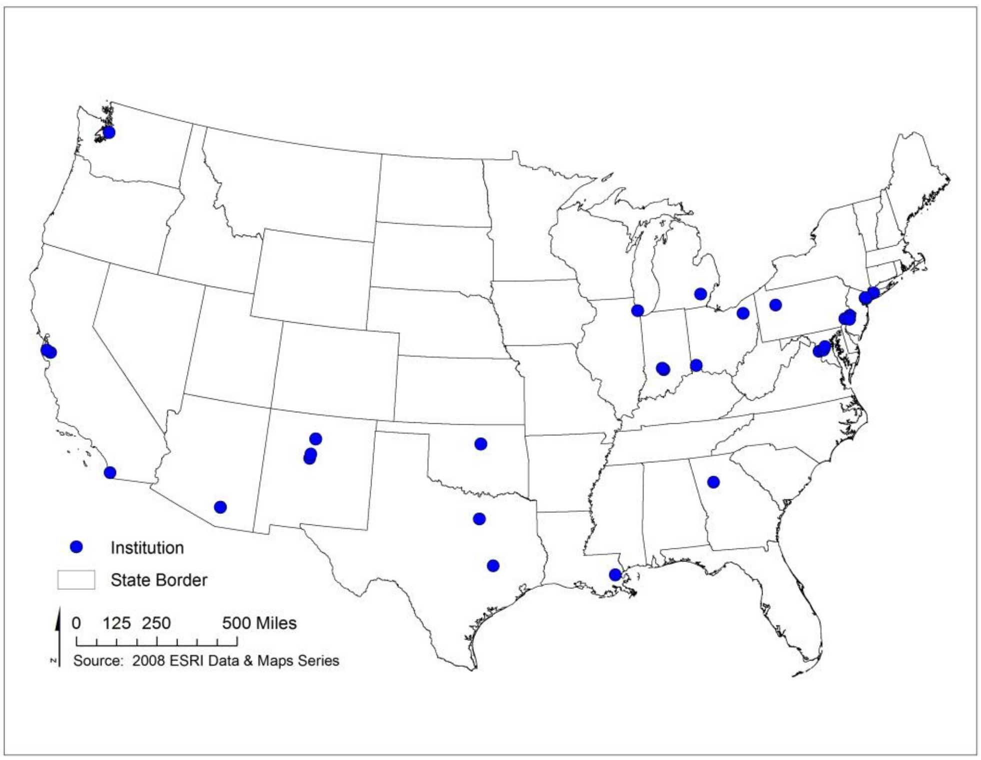

Figure 2. Institutional affiliation of cited publications.

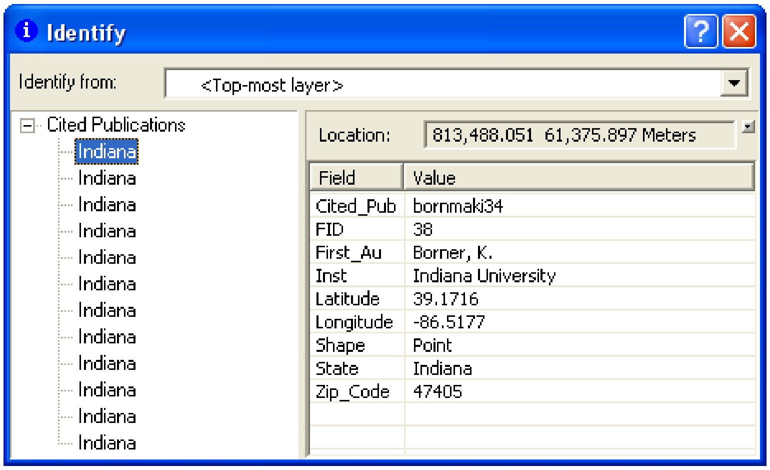

Figure 3. Cited publication fields.

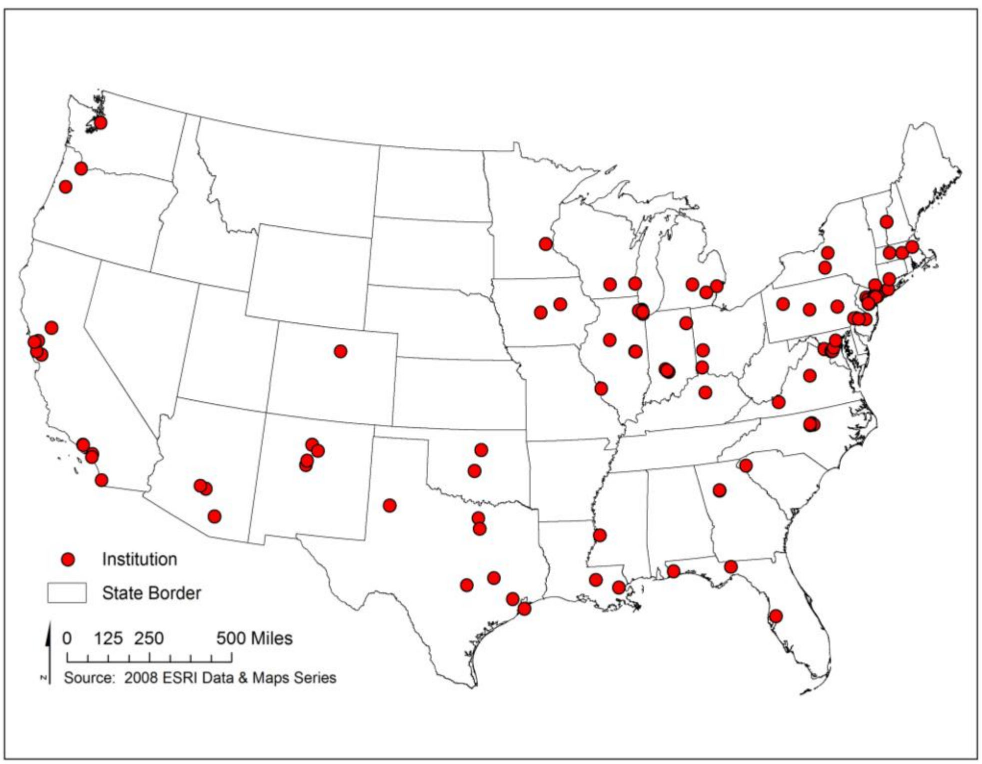

Figure 4. Institutional affiliation of citing publications.

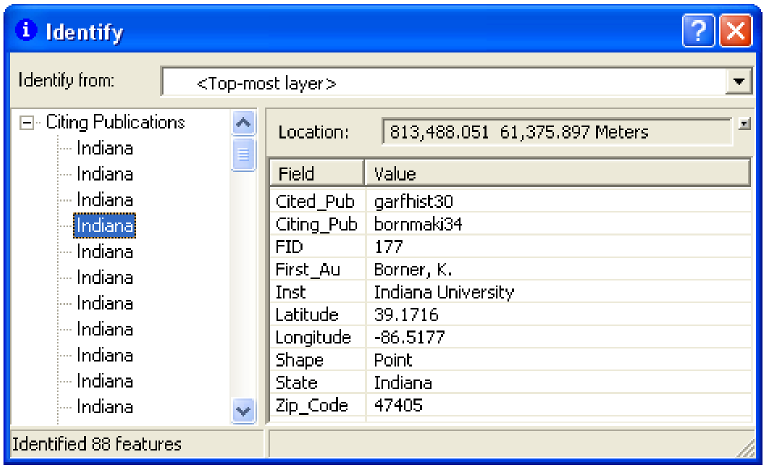

Figure 5. Citing publication fields.

The results of the average nearest neighbor analysis for cited and citing point maps are presented in Figure 6 and Figure 7, respectively. In both analyses, patterns of points were found to have a Z score ≤ -2.58 and therefore clustered with a 0.01 significance level. The point clustering is not apparent since many of the points were

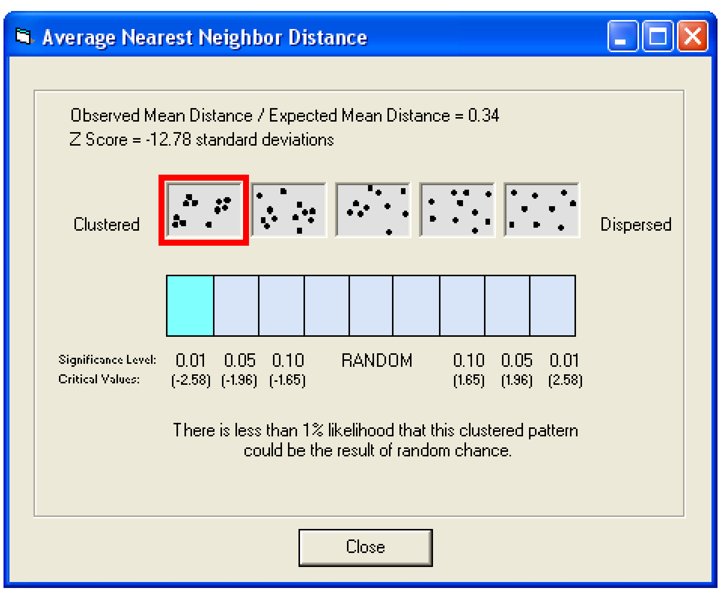

Figure 6. Average nearest neighbor analysis of cited publications.

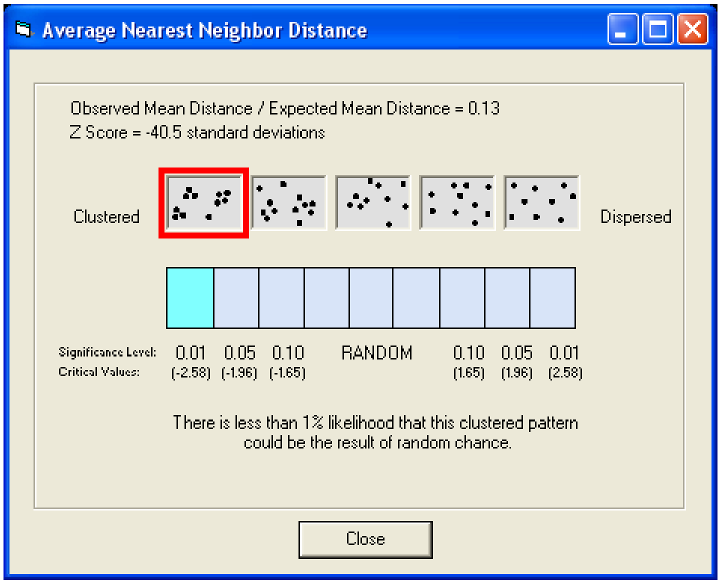

Figure 7. Average nearest neighbor analysis of citing publications.

associated with the same institution and superimposed.

Cumulative counts for both cited and citing publications were mapped using proportional bar symbols in Figure 8 and Figure 9, respectively. The clustering is much more apparent once proportional symbols rather than points are used. The fields associated with the cumulative number of cited and citing publications for each location are presented in Figure 10 and Figure 11, respectively. For example, Figure 10 and Figure 11 show the number of cited (10) and citing (79) publications for ZIP code 47405 (Indiana University) under the Cited and Citing fields, respectively. An alternative representation of the cumulative data, one more typical of bibliometric studies, is presented in Table 6 and shows the five institutions most frequently citing and being cited.

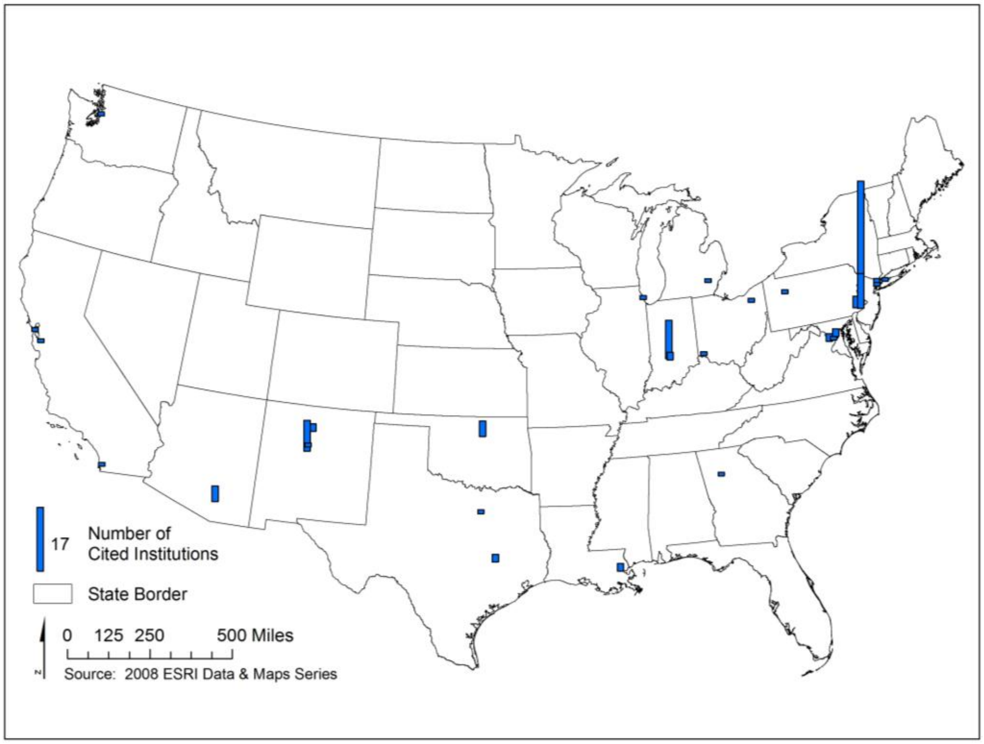

Figure 8. Institutional affiliation of cited publications – proportional symbol.

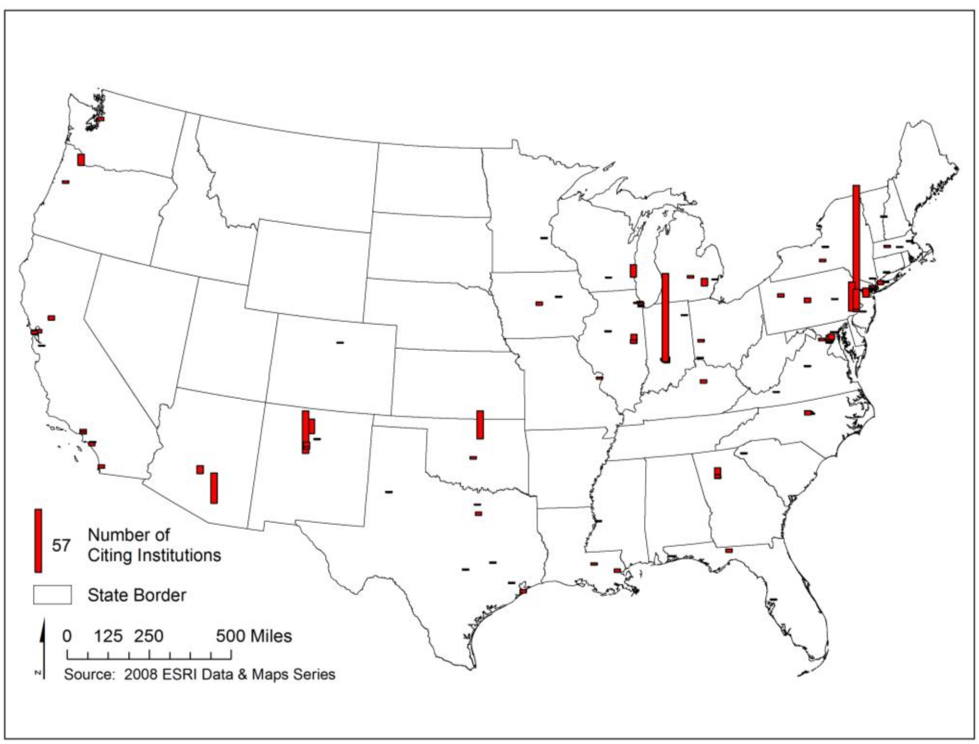

Figure 9. Institutional affiliation of citing publications – proportional symbol.

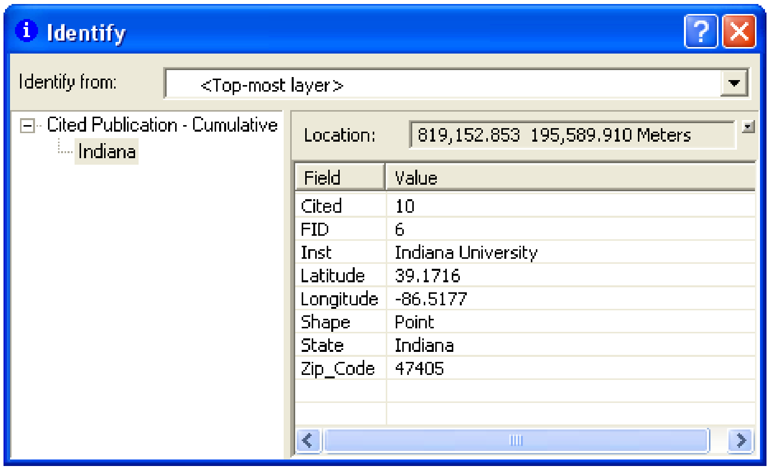

Figure 10. Cumulative number of cited publication fields.

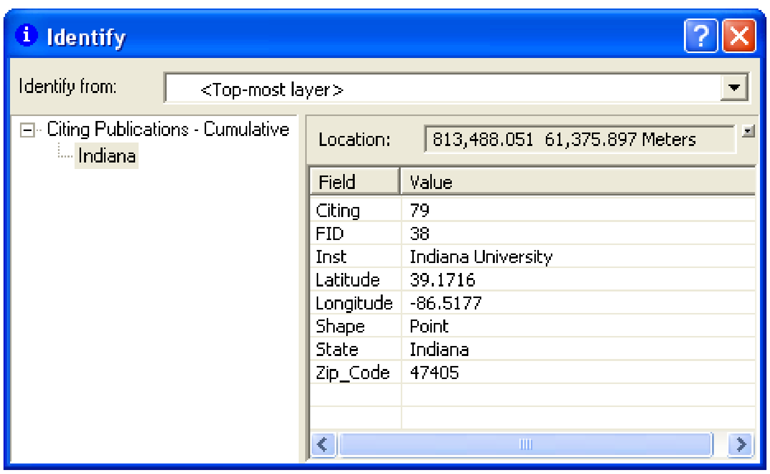

Figure 11. Cumulative number of citing publication fields.

Table 6. Top five cited and citing institutions.

	Cited Publications		Citing Publications	
Rank	Institution	Number	Institution	Number
1	Drexel University	33	Drexel University	113
2	Indiana University	10	Indiana University	79
3	Thomson ISI	9	Sandia National Lab	38
4	Sandia National Lab	8	University of Arizona	27
5	Oklahoma State University	4	SciTech Strategies Inc.	26
	University of Arizona	4		

4.3. Maps of Cited-Citing Publication Networks

A map showing the cited-citing publication networks for 1995-2009 is presented in Figure 12. Since none of the 102 original publications were cited during 1995 or 1996, no cited-citing network lines are shown for those two years.

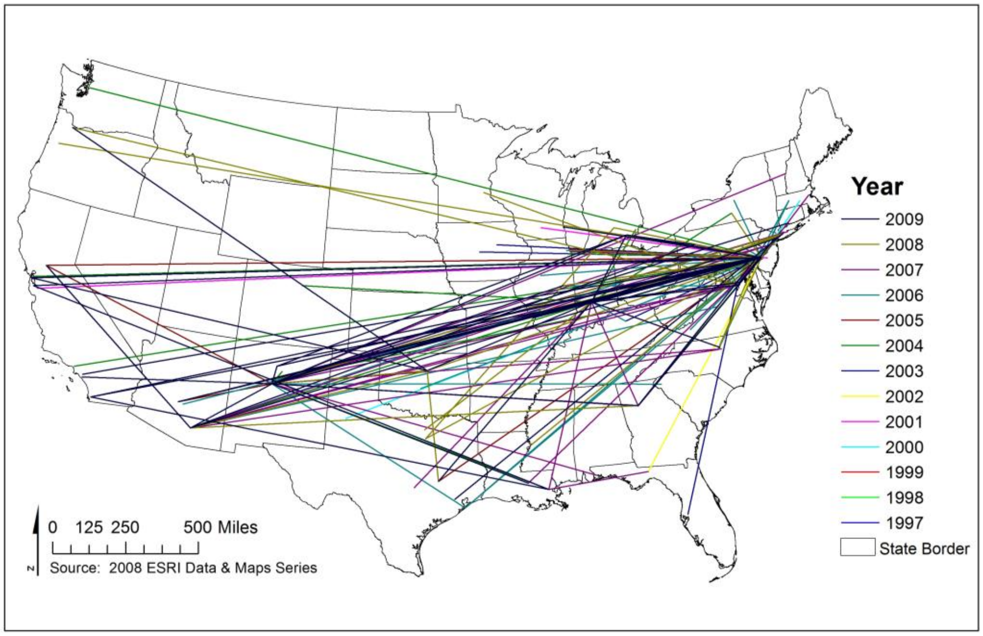

Figure 12. Cumulative cited-citing network, 1995-2009.

The map was created using a cited-citing layer for each of the 15 years and a contiguous United States base map. In cases where cited-citing publication network authors were located within the same ZIP code, no polyline was created. This occurred when the cited and citing authors were located at the same institution, including when authors cited their own publications (i.e., self-citation). This also occurred when cited and citing authors were located at two different institutions within the same ZIP code. Consequently, only 400 polylines were created from the 591 cited-citing publication networks. Of the 191 cited-citing publication networks that were not represented by a polyline, 165 cited-citing publication networks occurred between authors at the same institution and 134 of those were self-citations. The remaining 26 cited-citing publication networks without polylines resulted from authors located within a single ZIP code, but affiliated with different institutions.

Unique values (or hues) were assigned to cited-citing publication networks for each year; however, polylines for more recent years overlapped previous years in many instances. More specifically, 292 of the 400 polylines overlapped on the cumulative cited-citing publication network map. Using small multiples addressed this issue to some extent since each year is represented separately, but overlap occurred within the small multiples as well. Figure 13 presents a series of small multiples at 1-year intervals to illustrate each annual pattern of cited-citing networks over the 15-year period. An example of polyline fields for both the cumulative cited-citing network (Figure 12) and annual small multiples (Figure 13) is shown in Figure 14, which includes both the cited and citing publication pair.

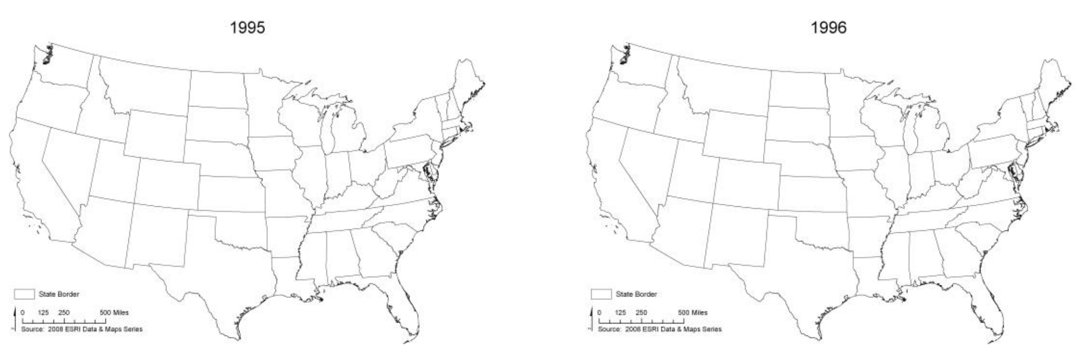

Figure 13. Small multiples of annual cited-citing networks, 1995-2009.

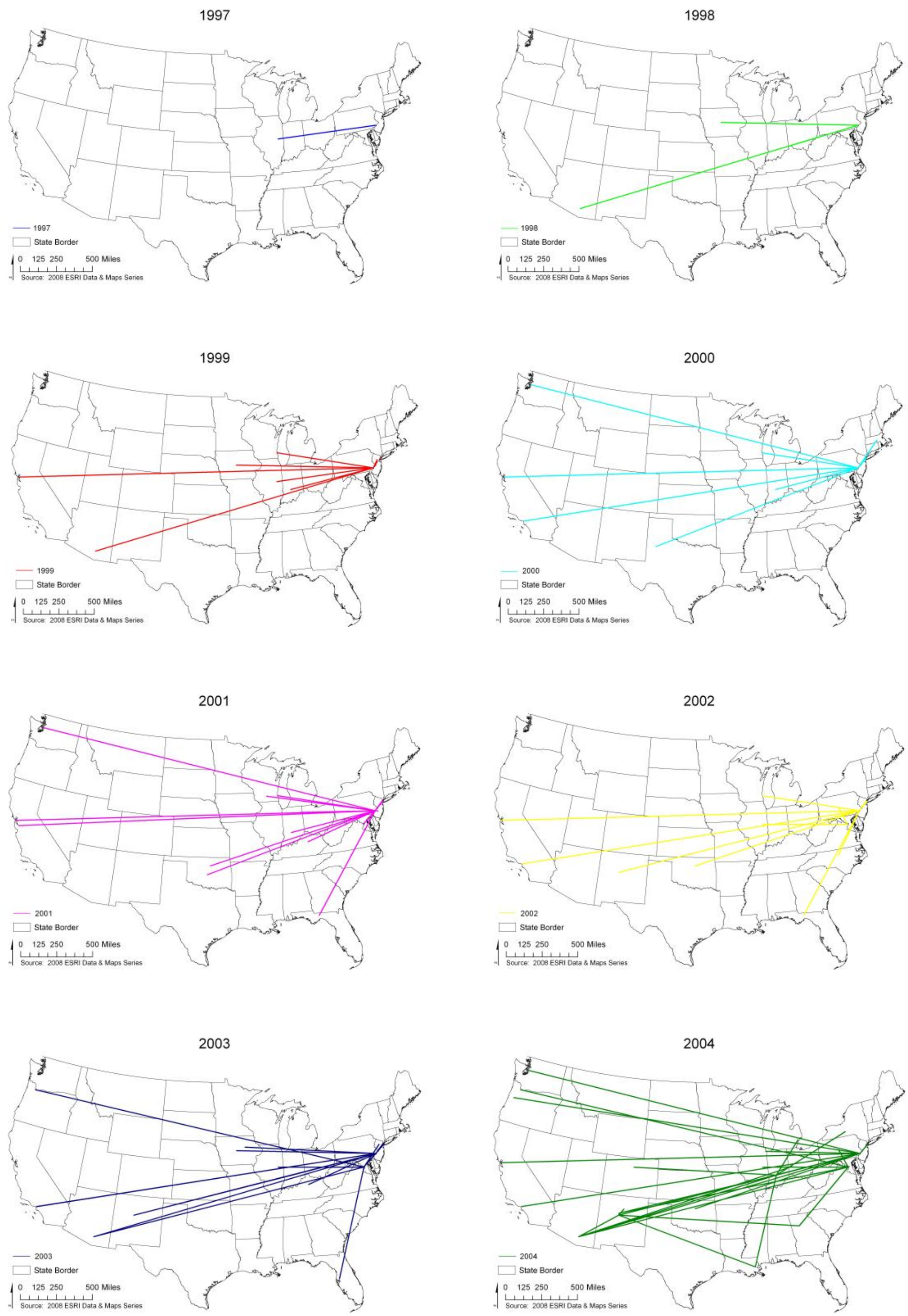

Figure 13. Small multiples of annual cited-citing networks, 1995-2009 (cont.).

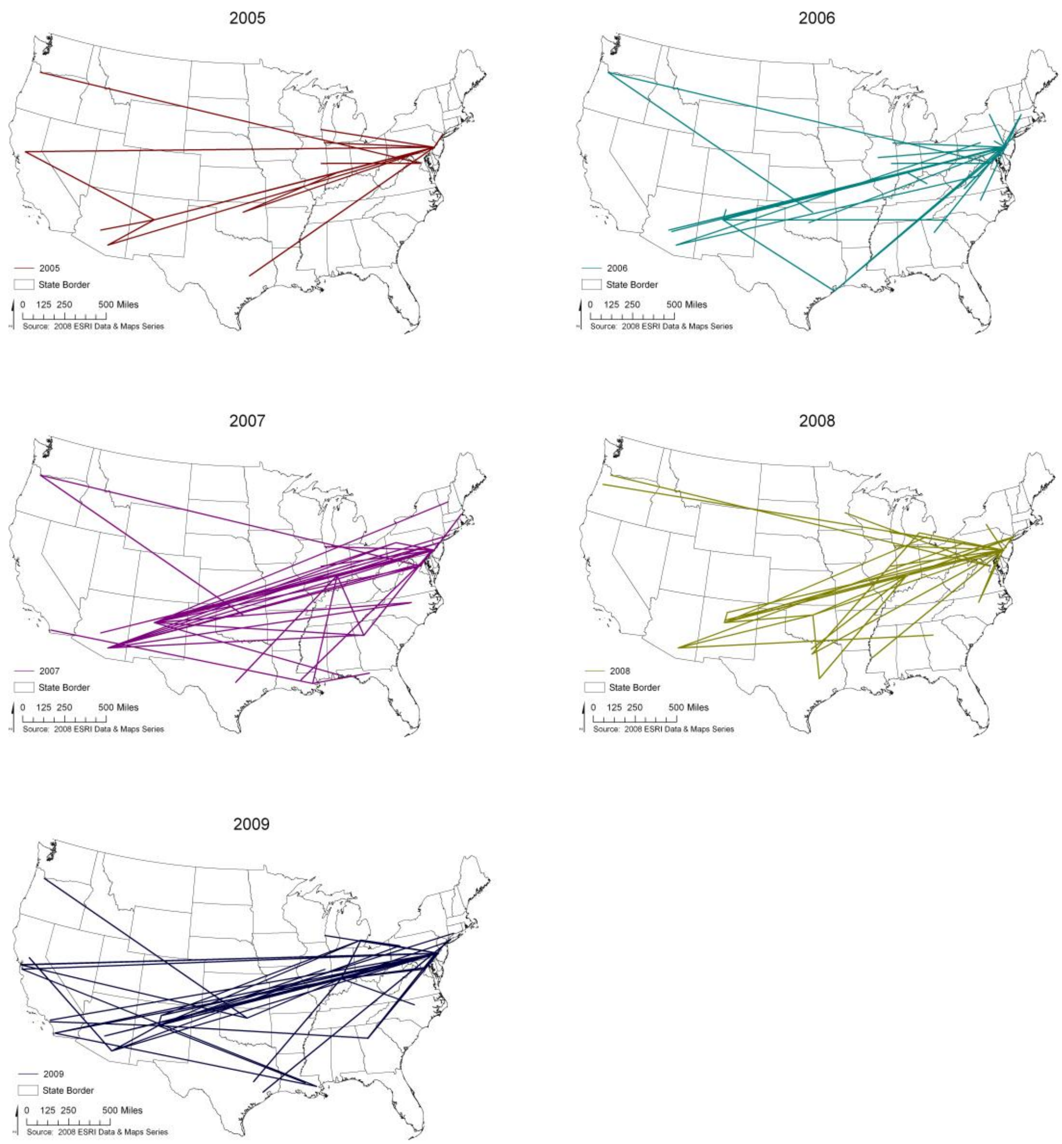

Figure 13. Small multiples of annual cited-citing networks, 1995-2009 (cont.).

A more thorough analysis of the cited-citing network patterns is presented in the Discussion section, but a few salient patterns are mentioned here. Most notable is the concentration of activity around Philadelphia, PA (Drexel University and Thomson ISI). It is not until 2004 that other centers of activity in the Southwest become apparent [Sandia, NM (Sandi National Lab) and Tucson, AZ (University of Arizona)]. Two other centers of activity, Bloomington, IN (Indiana University) and Stillwater, OK (Oklahoma State University) also emerge in 2004, but are less prominent due to overlapping cited-citing polylines. These general patterns correspond to the data presented in Table 6.

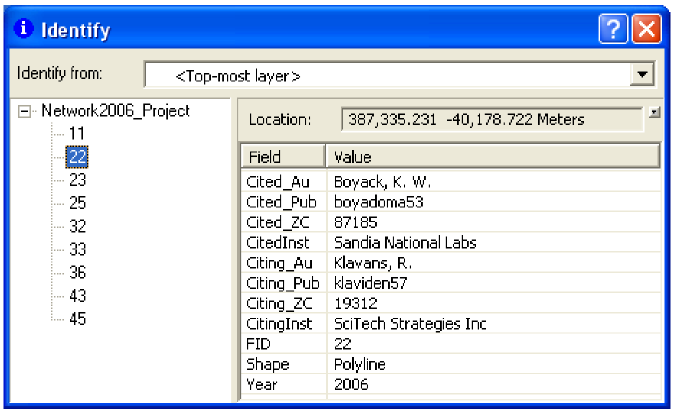

Figure 14. Cumulative cited-citing network fields.

Two additional sets of small multiples were created to show cumulative cited-citing publication networks at annual intervals. As indicated in the Conceptual Framework and Methodology section, several approaches were considered for each set of maps with respect to hue assignment. One of the approaches employed a single hue (blue). The result is shown in Figure 15 and the fields are identical to cited-citing publication networks shown in Figure 14. An alternative approach is to use the concept of advance-retreat, where the current cited-citing network is assigned red and older networks blue. Two 25-second animations were then created using the small multiples shown in Figure 15 and a similar set of small multiples using advance-retreat (not shown). The duration of the animations were set at 1.5 second per frame based on the research of Griffin *et al*. (2006). The effectiveness of animation compared to the small multiples is left to the user to decide, but it is argued that the two are complimentary rather than mutually exclusive. The animations are provided in Supporting Information as .wmv files.

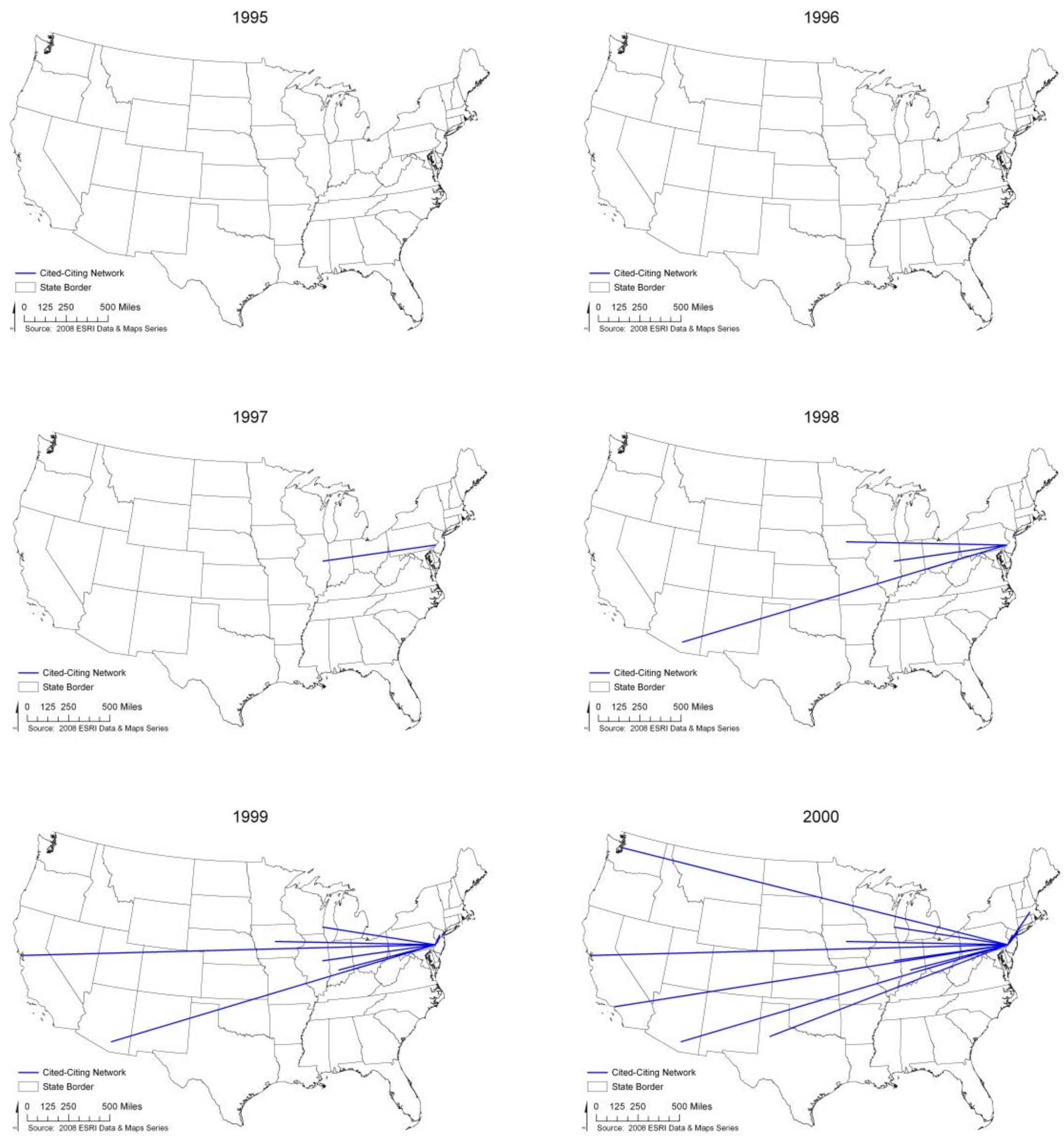

Figure 15. Cumulative small multiples of annual cited-citing networks, 1995-2009.

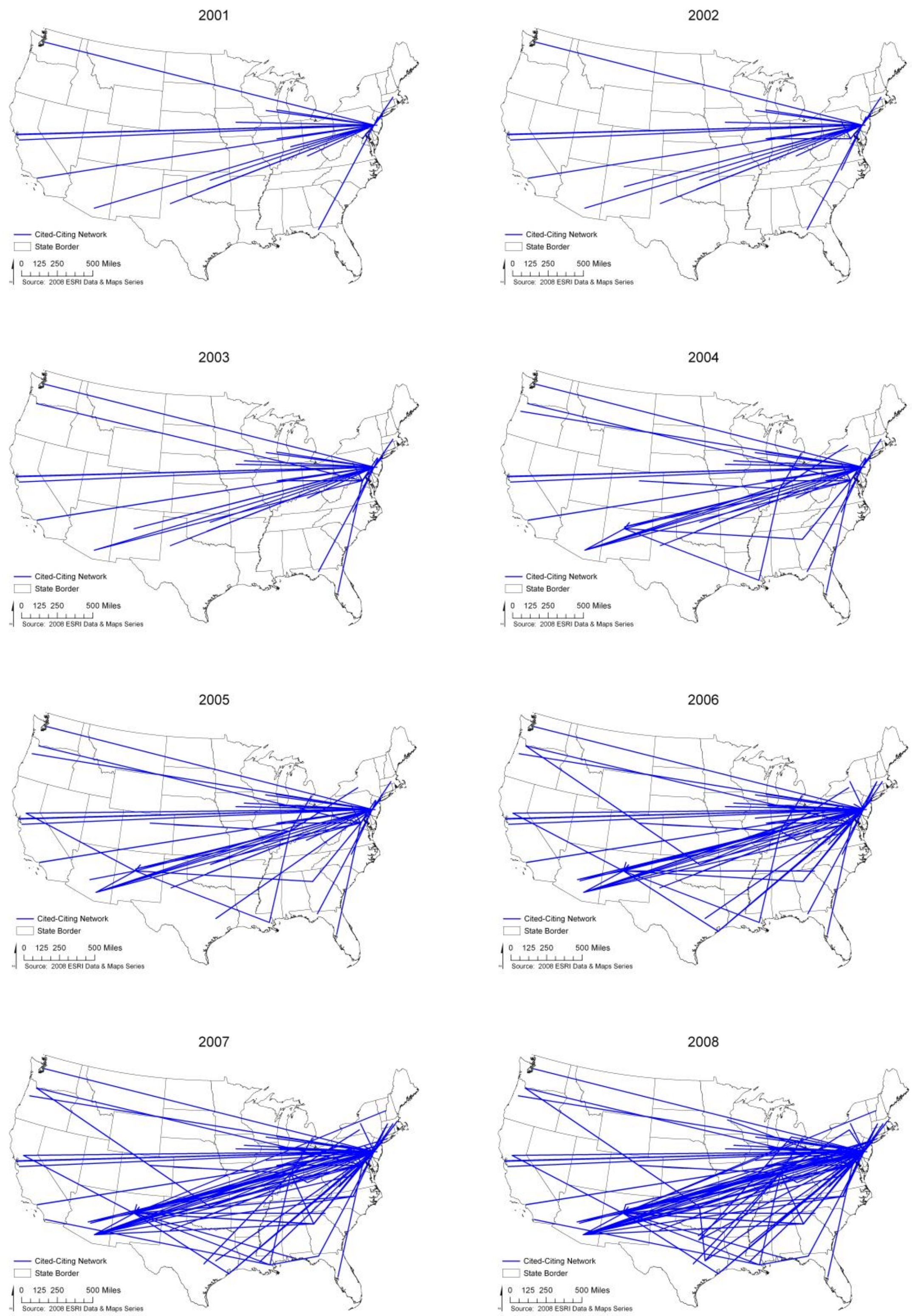

Figure 15. Cumulative small multiples of annual cited-citing networks, 1995-2009 (cont.).

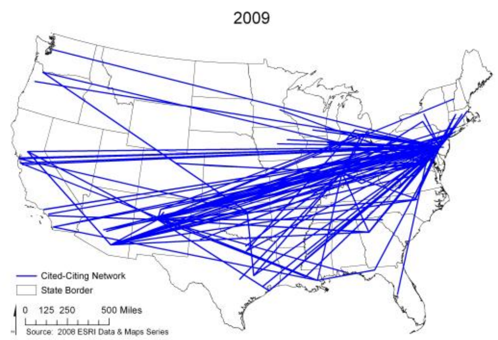

Figure 15. Cumulative small multiples of annual cited-citing networks, 1995-2009 (cont.).

4.4. Maps of Co-author Networks

The 102 cited publications had a total of 65 co-authored publications, but only 26 of those were co-authored by individuals located at different ZIP codes. The co-authors of the other 39 co-authored publications were located within the same ZIP code, including one instance of two institutions with the same ZIP code. Since no polylines resulted between co-authors within the same ZIP code, separate polyline layers were only created for the 26 co-author networks. Unlike the cumulative cited-citing network maps, the co-author network maps are shown in the same hue. This was done due to the difficulty in representing and differentiating 26 hues on the same map. The results of the cumulative co-author publication network are presented in Figure 16.

In addition to the cumulative co-author network map, small multiples were created to show the development of the co-author networks at annual intervals. Separate small multiples for each of the 26 co-author networks would have been overwhelming from a cognitive standpoint, so small multiples were created at annual intervals. The results were maps for 10 of the 15 years examined. In a situation similar to the cited-citing publication networks, overlapping occurred between co-author network polylines. The 26 co-author networks consisted of 33 individual polylines and 17 of those polylines overlapped with at least one other polyline. The unique values approach was used to assist with differentiating co-author networks within each year and dotted lines were employed where co-author networks overlapped. The results are presented in Figure 17 and the co-author networks are identified by their publication codes, which are listed in Appendix B. An example of the fields for the polylines in Figure 16 and Figure 17 is presented in Figure 18.

Figure 16. Cumulative co-author networks, 1995-2009.

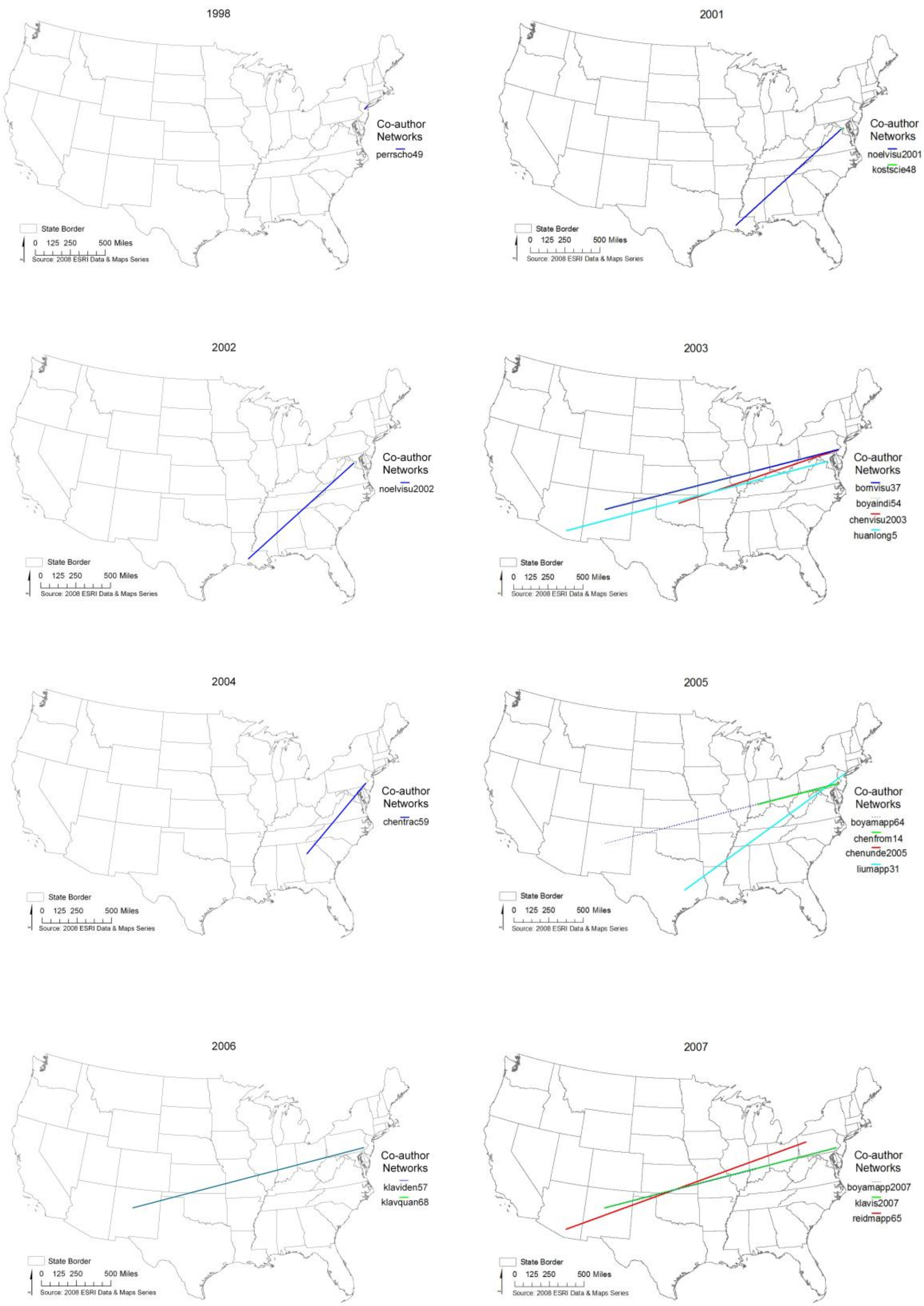

Figure 17. Small multiples of annual co-author networks, 1995-2009.

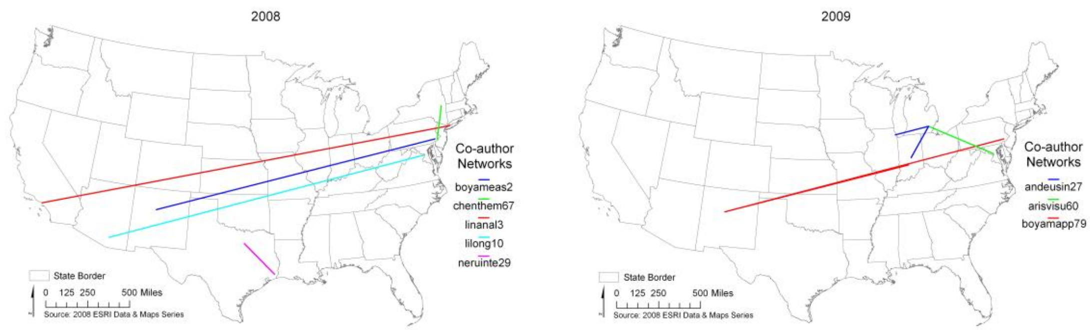

Figure 17. Small multiples of annual co-author networks, 1995-2009 (cont.).

The overall pattern is similar to the cited-citing publication network patterns (Figure 12, Figure 13, and Figure 15). The polylines show collaboration between authors in the Northeast and Southwest; however, the amount of collaboration is far less than the cited-citing patterns shown in the cited-citing publication networks.

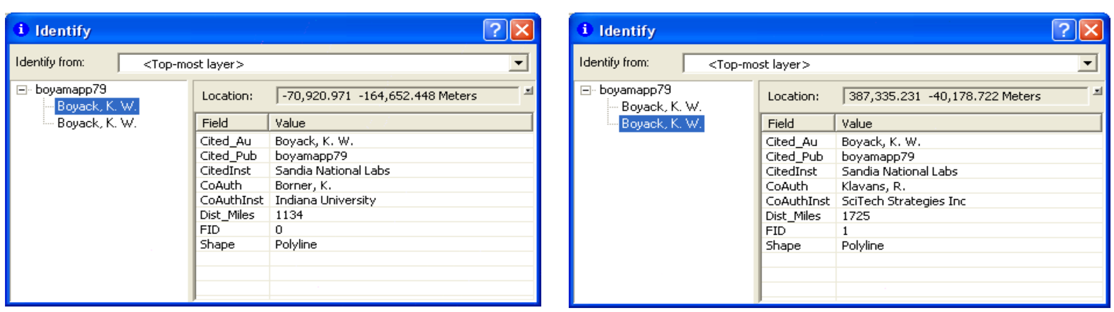

Figure 18. Co-author network fields.

4.5. Statistical Analysis of Co-author Distance

The average distance between co-authors was determined for all 65 co-authored publications and presented in Table 7. As indicated earlier, only 26 publications were co-authored by researchers at more than one institution or ZIP code. Consequently the distance between co-authors for the remaining 39 co-authored publications is zero;

meaning that all authors for the 39 publications were affiliated with the same institution or in one case resided in the same ZIP code.

Table 7. Average distance between co-authors (miles).

Year	Co-authored Publications											Annual Mean
	1	2	3	4	5	6	7	8	9	10	11	
1995	N/A											N/A
1996	N/A											N/A
1997	N/A											N/A
1998	29	0										14.5
1999	N/A											N/A
2000	0											0
2001	1024	13	0	0	0	0						172.8
2002	1024	0	0	0	0							204.8
2003	1213	1134	869	325	0	0						590.2
2004	662	0	0	0	0	0						110.3
2005	1433	1430	603	6	0	0	0	0	0			385.8
2006	1725	1725	0	0	0							690.0
2007	1843	1725	1430	0	0	0	0					714.0
2008	1718	496	388	129	67	0	0					399.7
2009	1430	281	210	0	0	0	0	0	0	0	0	174.6

The numbers of co-authored papers for 1995-2009 were also plotted and presented in Figure 19. The Pearson's correlation coefficient was determined to be 0.9009 with a p-value of < 0.0001. This indicated a statistically significant positive relationship between number of co-authored papers and year. More specifically, there is an increase in the number of co-authored publications from 1995 to 2009.

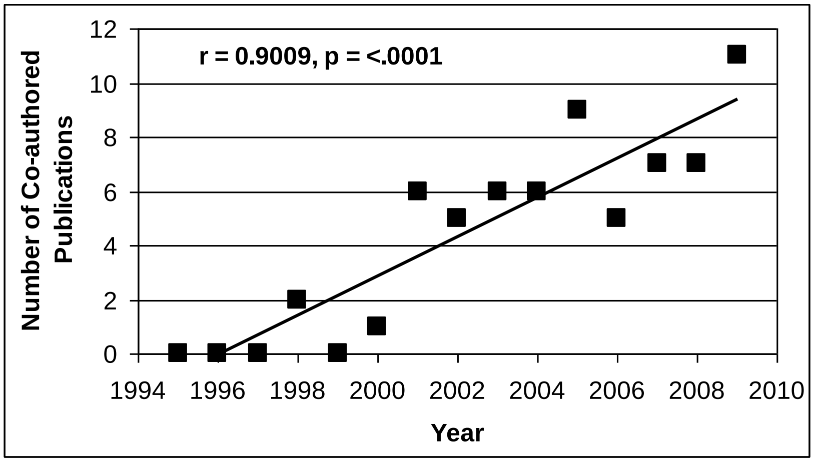

Figure 19. Number of co-authored publications, 1995-2009.

A visual inspection of Table 7 suggests that the average distance between co-authors within each year is not normally distributed. Since normality is an assumption for Analysis of Variance (ANOVA), the normality was tested using Shapiro-Wilk. The normality test was only conducted for 2001-2009, because there is insufficient data for 1995-2000 for both the normality test and any subsequent hypothesis testing. The p-values for each year (i.e., group or treatment) are presented in Table 8. Except for 2003, data for all years is not normally distributed at a p-value < 0.01.

Table 8. P-values from normality tests of average distance between co-authors.

Year	p-value
2001	<0.0001
2002	<0.0001
2003	0.1882
2004	<0.0001
2005	0.0004
2006	0.0065
2007	0.0066
2008	0.0050
2009	<0.0001

Since the data was not normally distributed, the following hypothesis was tested using Kruskal-Wallis, which is a nonparametric version of ANVOA:

H_0 = The annual mean distances are the same.

H_1 = The annual mean distances are different.

Where H_0 = Null Hypothesis and H_1 = Alternate Hypothesis

The Kruskal-Wallis test resulted in a p-value of 0.5205, so the null hypothesis is not rejected. Consequently there is no difference in annual mean distances from 2001 to 2009. A more detailed SAS output is presented in the Appendix C.

4.6. Discussion

The search of Web of Science utilized a single, yet complex search string to retrieve publications related to visualization of bibliographic data. However, keyword searching is known to be imperfect and will not capture all relevant publications due to variations in terminology and variant spellings. The subsequent refining within Web of Science (i.e., document type and publication year), relevancy of each publication, and the need to restrict co-authors to those residing in the contiguous United States reduced the size of the data set. The criteria used to determine the relevancy of each publication was applied consistently (i.e., each contained a visualization of bibliographic data versus tables or standard graphs); however, there is always some subjectivity associated with this type of categorization. Ensuring that co-authors were all residing in the contiguous United States was crucial to analysis of co-author distances. The level of international collaboration was also underestimated, and this also limited the size of the data set. These limitations do not undermine the results, but they do indicate the data mapped is more a sample rather than a definitive data set of all published works on the visualization of bibliographic data.

The 2010 ZIP code centroid was used as a surrogate for institutions and all data were mapped using that centroid. While using institutional affiliation based on the first author is standard practice in bibliometric studies, it is important to note that the ZIP code centroid does change over time. While this approach appears to introduce errors, the distances involved compared to the changes in the centroid would be small. An alternative approach is to geocode the institution itself since it does not change, but this approach has the potential to introduce similar errors and some challenges. These include looking up and determining what address to geocode (e.g., main mailing address, departmental address, etc.) and while these addresses generally remain constant, mailing addresses do change over time.

Several approaches were employed to assign a hue to the cited-citing and co-author network maps. These approaches included advance-retreat, single hue, and unique values method. Some of these appear to be more effective than others for this study, but that is ultimately decided by the user. The cumulative cited-citing network map and annual small multiples were created with the unique values method. Comparing the cumulative cited-citing network map (Figure 12) to the 2009 annual cumulative small multiple (Figure 15) containing the same cited-citing networks, the unique values method appears to be less effective than the single hue. In the opinion of this author, the use of additional hues causes more confusion than clarity. While it may be possible to identify individual networks, overlapping polylines and additional hues make it more difficult to visualize overall patterns. In this case, a single hue may be more effective.

The growth of the cited-citing network is clearly shown in the small multiples (Figure 15), which allows users to compare and explore the time-series at their own pace for patterns and change. The small multiples reveal several notable trends. The initial citing originates mostly from Pennsylvania with other locations increasing over time. The increased activity can be seen as a “fanning feature” as polylines emanate from a location. There appears to be a step increase in 2004 with respect to citing behavior from different geographic locations generally. Another noticeable change occurs in 2007 when researchers in New Mexico and Arizona are increasingly being cited and citing the work of others.

The cited-citing network was also presented using animation, which provided a more dynamic visualization of the data. One animation was created using the cumulative annual cited-citing network maps (Figure 15) and another using cumulative annual cited-citing small multiples where an advance-retreat approach was employed. In addition to the visual variables utilized in the small multiples, display date was crucial in communicating when change occurred and to a lesser degree the order. In the case of the advance-retreat animation, temporal changes were also communicated with hue changes with the current year displayed in red and previous years in blue. Unlike the static small multiples, changing hues and display date serve as a type of “blinking” that draws attention and communicates when change occurs.

Since the magnitude of change was governed by the data itself and the frequency was fixed (i.e., one frame per year), selecting the proper duration had a major impact on the rate of change and ultimately the smoothness of the transition. If the duration is too long, the animation would appear choppy. If the duration is too short the viewer will not have sufficient time to comprehend the changes. The duration was ultimately set at 1.5 seconds, a value utilized by Griffin *et al.* (2006) in their cluster map animations. The selection of 1.5 seconds for the duration appears to provide the proper balance between enough time to view each frame and smoothness; however, user testing is required to determine if 1.5 seconds is ideal in terms of user cognition and transition. The same is true of the effectiveness of the animations compared to small multiples. It is the opinion of this author that the two approaches are complimentary with each being effective in different ways. Small multiples provide a user more time to study each year in more depth and to compare years in a nonsequencial manner. The animations could be paused to study each year in more depth, but their strengths reside in seeing a holistic presentation of the data at a constant rate. While the growth of cited-citing networks (i.e., initially Philadelphia, PA and then later New Mexico and Arizona) is apparent in the small multiples, it is the opinion of this author that it is more pronounced in the animation than in the small multiples. The use of the advance-retreat approach appears to be somewhat effective in presenting current changes, but it may distract the user from seeing overall patterns.

In both the small multiples and animations, cited and citing publication networks of Indiana University are lost among the other cited-citing polylines. Indiana University publications were heavily cited and were citing numerous publications (Figure 8 and Figure 9), but since Bloomington, Indiana resides almost perfectly on a line drawn between Philadelphia, PA and Sandia, NM, it is completely lost in both approaches. The same is true of locations that frequently cite the same publications or were cited by the same individuals since the polylines are superimposed (e.g., Stillwater, OK). The last situation was somewhat mitigated using the changing hues to draw attention to the more current networks. While this appears to understate the level of activity shown using the proportional symbol maps, it is important to note that all maps are a representation of a phenomenon and that often more than one map (or type of map) is needed to represent that reality. Using all three maps (i.e., proportional symbol maps, cited-citing maps, and animation) overcomes some of the limitations of any single map in this study.

As noted in the Analysis Results and Discussion section, there were only 65 co-authored publications among the 102 publications. Of the 65, only 26 co-authored publications were co-authored by individuals at more than one institution. The remaining 39 co-authored publications were authored by individuals within the same ZIP code. Like the cited-citing network maps, co-author networks also have overlapping polylines. This was especially true of co-authors that regularly collaborated. The use of unique colors assisted in distinguishing the co-author relationships, but some were still obscured. Consideration was given to presenting the 26 co-author networks as individual small multiples to provide more distinction between co-author networks, but this would detract from the cumulative impact seen in Figure 16 and Figure 17. One of those cumulative impacts is the similarity to some of the cited-citing network small multiples. Based on visual inspection of the pattern, some of the co-author networks appear to be the same as the cited-citing networks. More specifically, the co-author networks between Philadelphia, PA and the Southwestern United States look similar.

The distance between each 26 co-authored publications was calculated using the Spatial Statistics Tools in the ArcToolbox. The polylines were created from the first author to each of the co-authors. So the distances calculated had the first author as the primary node, which did not necessarily minimize the co-author network distance. However, this model makes sense in that communications are typically routed through the first author.

The cumulative distances for each of the 65 co-authored publications tended to be of extremes. The co-authors were either at the same institution or some large distance away. As indicated in the results, there were less than three co-authored publications from 1995-2000. This was a period of fewer total articles on this topic and more single authored articles. While fewer total articles are not surprising, the larger number of single authors is interesting to note. There proved to be a statistically significant increase in the number of co-authored publications, which may be explained by the fact that this field was in its infancy during the mid-1990s and few collaborators existed. As the field matured, collaboration increased.

The dearth of data during the earliest years (i.e., 1995-2000) resulted in limiting the hypothesis testing to 2001-2009. While statistical valid in terms of the size of the data set, additional data may make a more convincing argument and interesting visualizations. The mid-1990s is a period that experienced the most rapid improvements in telecommunications, so it was unfortunate that the period had insufficient data to test.

Chapter 5

Conclusion

5.1. Summary

This study demonstrates that the spatial aspects of bibliographic data can be represented using ArcGIS as both points and polyline networks, though there are some limitations to the model and visualizations. Comparing this study to those in relative space, mapping bibliographic data geographically provides opportunities to explore spatial patterns that give insights into the relationships of authors, co-authors, and the research. For example, average nearest neighbor analyses showed that both cited and citing institutions were clustered. ArcGIS provided a clear visualization of that geographic clustering and a means to test it statistically, something not performed in relative space or discussed in previous studies involving visualization of bibliographic data using GIS. While not particularly unique to ArcGIS, the software provides tools to compute co-author distance and ultimately test the "death of distance" hypothesis statistically. The annual mean distances between co-authors were determined to be the same for 2001-2009. Based on the average annual distance between co-authors, the "death of distance" did not occur in visualization of bibliographic data among co-authors in the United States. However, this is a new field and 60% of the collaborations studied involved co-authors at the same institutions. One should be cautious about generalizing these findings to other fields or geographic extents, but this study contributes to a small body of research that challenges "death of distance" involving knowledge creation and diffusion.

The results of this study are consistent with two studies examining geographic distances between co-authors (Hoekman *et al*. 2010, Maggioni and Uberti 2009), which specifically challenged the “death of distance” hypothesis and found collaboration to be localized. However, Adams *et al*. (2005) and Waltman *et al*. (2011) found increasing co-author distances. This apparent discrepancy may be due to differing collaboration patterns among disciplines or geographic extents studied. In contrast to the current study, all four of these previous studies were highly aggregated (i.e., broad disciplines). To infer that these highly aggregated studies describe co-authorship within a particular narrowly defined field could result in an ecological fallacy, hence the need to study discrete fields. Beyond the specific findings for the visualization of bibliographic data, this study advances the field by demonstrating how ArcGIS can be used to visualize cited-citing publication networks and determine spatial trends within a field. It is also the first study to utilize spatial analyses within ArcGIS to explore these geographic relationships among co-authors.

5.2. Limitations

Several limitations were mentioned throughout this study. These include the size of the data set, overlapping cited-citing and co-author network polylines, and utilization of the 2010 ZIP code centroid for all years. There were also numerous cited-citing publication and co-author networks that resulted in no polylines since the authors were located within the same ZIP code. While the overall cited-citing publication and co-author network patterns were representative of networks between authors and co-authors in different ZIP codes, networks between authors and co-authors within the same ZIP code were not represented. Consequently, the level of local collaboration (i.e., same institution) and self-citation were not adequately represented. There were also limitations associated with the conceptual model, which included equating one ZIP code to one institution, limiting publications to authors residing in the United States, and determining co-author distance based on first author.

5.3. Further Research

There are several avenues for future research. User studies are needed to investigate the effectiveness of the visualizations of bibliographic data. This includes the overall representation, as well as comparisons between small multiples, animations, and use of hue. Processes in this study were labor intensive and this would be challenging to scale. Therefore, new workflows should be developed to streamline importation of data from bibliographic databases, especially the geographic aspects, directly into ArcGIS.

A number of the limitations outlined in Section 5.2 could be addressed in future research. For example, this study focused on a narrow area of bibliometrics (i.e., visualization of bibliographic data), which resulted in a small sample size. Future studies could increase the sample size by broadening the field of study to bibliometrics, a field with a longer history and comprised of more publications, or pursue an entirely different field that is more widely studied such as visualization. Beyond studying other fields or entire disciplines, the approach outlined in this study could be applied to exploring clustering and collaboration within specific industries or explore technological trends using the patent literature (e.g., nanotechnology, cellular phones, etc.).

Addressing the overlapping cited-citing publication and co-author network polylines provides another opportunity. One possible solution is to use a single polyline between ZIP codes and scale the thickness of polylines to the number of network connections between ZIP codes. The scaled polylines could also be displayed using transparency to facilitate display of partial overlaps and intersecting polylines.

Another opportunity for improvement is to address the absence of polylines when cited-citing publication and co-author network authors were located within the same ZIP code (e.g., local collaboration or self-citation). One possible solution is to add proportional symbols at each node (i.e., ZIP code) to represent all cited-citing publication or co-author networks. This would communicate the total number of networks regardless of whether a polyline was created or not.

Appendix A

Create Polyline Shapefile from Table Sample

The following Visual Basic for Applications (VBA) script takes two X,Y coordinate pairs and creates a polyline shapefile from a dBASE file. Instructions for using the VBA script follow:

1. Open ArcCatalog
2. Paste VBA script below into Visual Basic Editor in ArcCatalog
   [Tools -> Macros -> Visual Basic Editor]
3. Navigate and select a dBASE file containing X,Y coordinate pairs
4. Click “Run” along toolbar in Visual Basic Editor
5. Rename the newly created shapefile (LineFC)
6. Repeat for each year and group of co-authors

NOTE: The following VBA script was developed for ArcGIS 9.0 or higher. In addition to the two X,Y coordinate pairs, the dBASE file must contain two columns labeled ID and LABEL. The two additional columns may remain empty.

The VBA script was obtained from the ESRI Developer Network (ESRI, 2004) and has been placed in the public domain with the following copyright statement and disclaimer.


```
Private Const m_sX1 As String = "X1"
Private Const m_sY1 As String = "Y1"
Private Const m_sX2 As String = "X2"
```

```
Private Const m_sY2 As String = "Y2"
Private Const m_sAttrib1 As String = "ID"
Private Const m_sAttrib2 As String = "Label"
Public Sub MakeLineFC()
  On Error GoTo ErrorHandler

  If Not TypeOf Application Is IGxApplication Then Exit Sub

  Dim pApp As IGxApplication
  Set pApp = Application

  If pApp.SelectedObject.Category = "dBASE Table" And TypeOf pApp.SelectedObject Is IGxDataset
Then
    Dim pGxDataset As IGxDataset
    Set pGxDataset = pApp.SelectedObject

    If pGxDataset.Type = esriDTTable Then

      '  Get the Selected Table.
      Dim pDataset As IDataset, pTable As ITable
      Set pDataset = pGxDataset.Dataset
      Set pTable = pGxDataset.Dataset

      '  Find Fields containing X and Y coordinates, and the specified attributes.
      Dim l_X1 As Long, l_Y1 As Long, l_X2 As Long, l_Y2 As Long, l_A1 As Long, l_A2 As Long
      l_X1 = pTable.FindField(m_sX1)
      l_Y1 = pTable.FindField(m_sY1)
      l_X2 = pTable.FindField(m_sX2)
      l_Y2 = pTable.FindField(m_sY2)
      l_A1 = pTable.FindField(m_sAttrib1)
      l_A2 = pTable.FindField(m_sAttrib2)
      If (l_X1 < 0) Or (l_Y1 < 0) Or (l_X2 < 0) Or (l_Y2 < 0) Or (l_A1 < 0) Or (l_A2 < 0) Then
        MsgBox "Could not find specified Fields"
        Exit Sub
      End If

      '  Set up a Fields collection for the new Feature Class.
      Dim pField As esriGeoDatabase.IField, pFieldEdit As esriGeoDatabase.IFieldEdit
      Dim pFields As esriGeoDatabase.IFields, pFieldsEdit As esriGeoDatabase.IFieldsEdit
      Dim pGeomDefEdit As IGeometryDefEdit, pSR As ISpatialReference

      Set pFields = New esriGeoDatabase.Fields
      Set pFieldsEdit = pFields
      pFieldsEdit.FieldCount = 3

      ' Create the geometry field.
      Set pGeomDefEdit = New GeometryDef
      Set pSR = New esriGeometry.UnknownCoordinateSystem
      With pGeomDefEdit
        .GeometryType = esriGeometryPolyline
        .HasM = False
        .HasZ = False
        Set .SpatialReference = pSR
      End With

      Set pFieldEdit = New Field
```

```
With pFieldEdit
  .Name = "Shape"
  .AliasName = "Geometry"
  .Type = esriFieldTypeGeometry
  Set .GeometryDef = pGeomDefEdit
End With
Set pFieldsEdit.Field(0) = pFieldEdit

'  Set the two attribute Fields by cloning from the existing Table.
Dim pClone As IClone
Set pClone = pTable.Fields.Field(l_A1)
Set pFieldsEdit.Field(1) = pClone.Clone
Set pClone = pTable.Fields.Field(l_A2)
Set pFieldsEdit.Field(2) = pClone.Clone

'  Now create the new Shapefile. First create a Feature UID.
Dim pCLSID As esriSystem.UID
Set pCLSID = New UID
pCLSID.Value = "esriGeoDatabase.Feature"

'  Now create a new shapefile FeatureClass (check the file does not exist first).
Dim pFSO As Object, sFCName As String
sFCName = pDataset.Workspace.PathName & "\LineFC.shp"
Set pFSO = CreateObject("Scripting.FileSystemObject")
If pFSO.FileExists(sFCName) Then
  MsgBox "Select different name for the new shapefile", vbInformation, "File of same name exists"
  Exit Sub
End If
Dim pFeatClass As IFeatureClass, pWksp As IWorkspace, pFeatWksp As IFeatureWorkspace,
   pWkspFact As IWorkspaceFactory
Set pWkspFact = New ShapefileWorkspaceFactory
Set pFeatWksp = pWkspFact.OpenFromFile(pDataset.Workspace.PathName, 0)
Set pFeatClass = pFeatWksp.CreateFeatureClass("LineFC", pFields, pCLSID, Nothing, esriFTSimple,
   "Shape", "")

'  Now, create the Line data and add it to the new FeatureClass along with the  specified attributes.

If pFeatClass Is Nothing Then Exit Sub
Dim l_FCA1 As Long, l_FCA2 As Long
l_FCA1 = pFeatClass.FindField(m_sAttrib1)
l_FCA2 = pFeatClass.FindField(m_sAttrib2)

'  Iterate all the rows in the selected Table.
Dim pTableCursor As ICursor, pRow As IRow
Set pTableCursor = pTable.Search(Nothing, True)
If pTableCursor Is Nothing Then Exit Sub
Set pRow = pTableCursor.NextRow

Dim pGeomColl As IGeometryCollection, pSegColl As ISegmentCollection
Dim pLine As ILine, pPolyline As IPolyline, pFeat As IFeature

Do While Not pRow Is Nothing

  ' For each row in the Table, create a PolyLine.
  Set pLine = CreateLn(CreatePt(pRow.Value(l_X1), pRow.Value(l_Y1)), CreatePt(pRow.Value(l_X2),
    pRow.Value(l_Y2)))
```

```
        Set pSegColl = New Path
        pSegColl.AddSegment pLine
        Set pGeomColl = New Polyline
        pGeomColl.AddGeometry pSegColl
        Set pFeat = pFeatClass.CreateFeature
        Set pPolyline = pGeomColl

        '  Set the Feature's Shape and the specified attributes.
        Set pFeat.Shape = pPolyline
        pFeat.Value(l_FCA1) = pRow.Value(l_A1)
        pFeat.Value(l_FCA2) = pRow.Value(l_A2)
        pFeat.Store

        Set pRow = pTableCursor.NextRow
      Loop
    End If
  End If

  'Refresh parent to show newly created file
  Dim pGxObject As IGxObject
  Set pGxObject = pGxDataset
  pGxObject.Parent.Refresh

Exit Sub

ErrorHandler:
  If Err.Number <> 0 Then
    MsgBox Err.Description, vbCritical, "Error: " & Err.Number
  End If
End Sub

Private Function CreatePt(ByVal dX As Double, ByVal dY As Double) As IPoint
  Set CreatePt = New Point
  CreatePt.PutCoords dX, dY
End Function

Private Function CreateLn(ByRef pPointFrom As IPoint, pPointTo As IPoint) As ILine
  Set CreateLn = New esriGeometry.Line
  CreateLn.PutCoords pPointFrom, pPointTo
End Function
```

Appendix B

Cited Publications and Publication Codes

Publication Code	Cited Publication	Number of Co-authors	Citation Count
alleadap2005	Allendoerfer, K., *et al.*, 2005. Adapting the cognitive walkthrough method to assess the usability of a knowledge domain visualization. *IEEE Symposium on Information Visualization*, 23-25 October Minneapolis. Piscataway, NJ: IEEE, 195-202.	7	0
andeusin27	Anderson, C.A., Keenan, G., and Jones, J., 2009. Using bibliometrics to support your selection of a nursing terminology set. *CIN-Computers Informatics Nursing*, 27 (2), 82-90.	3	1
arisvisu60	Aris, A., Shneiderman, B., Qazvinian, V., and Radev, D., 2009. Visual overviews for discovering key papers and influences across research fronts. *Journal of the American Society for Information Science and Technology*, 60 (11), 2219-2228.	4	0
barnmapp54	Barnhurst, K.G., Vari, M., and Rodriguez, I., 2004. Mapping visual studies in communication. *Journal of Communication*, 54 (4), 616-644.	3	3
bergaugm2009	Bergstrom, P. and Atkinson, D.C., 2009. Augmenting the exploration of digital libraries with web-based visualizations. *Fourth International Conference on Digital Information Management*, 1-4 November 2009 Ann Arbor, MI. New York: IEEE.	2	0
blatdiff80	Blatt, E.M., 2009. Differentiating, describing, and visualizing scientific space: a novel approach to the analysis of published scientific abstracts. *Scientometrics*, 80 (2), 385-406.	1	0
bollmapp69	Bollen, J. and De Sompel, H.V., 2006. Mapping the structure of science through usage. *Scientometrics*, 69 (2), 227-258.	2	6
bollclic4	Bollen, J., *et al.*, 2009. Clickstream data yields high-resolution maps of science. *PLoS One*, 4 (3), 4803.	7	4
bornvisi3960	Borner, K., 2000. Visible threads: a smart VR interface to digital libraries. *In:* R.F. Erbacher, *et al.*, eds. 7th Meeting on Visual Data Exploration and Analysis (v. 3960), 24-26 January 2000 San Jose, CA. Bellingham, WA: SPIE, 228-237.	1	0
bornmaki34	Borner, K., 2007. Making sense of mankind's scholarly knowledge and expertise: collecting, interlinking, and organizing what we know and different approaches to mapping (network) science. *Environment and Planning B-Planning & Design*, 34 (5), 808-825.	1	0
bornvisu37	Borner, K., Chen, C.M., and Boyack, K.W., 2003. Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37, 179-255.	3	105

bornrete2009	Borner, K., *et al.*, 2009. Rete-Netzwerk-Red: analyzing and visualizing scholarly networks using the network workbench tool *In:* B.Larsen and J. Leta, eds. *12th International Conference of the International Society for Scientometrics and Informetrics,* 14-17 July 2009 Rio de Janeiro. Leuven: ISSI, 619-630.	9	0
bornspat2005	Borner, K. and Penumarthy, S., 2005. Spatio-temporal information production and consumption of major US research institutions. *10th International Conference of the International Society for Scientometrics and Informetrics,* 24-28 July 2005 Stockholm. Stockholm: Karolinska University Press, 635-641.	2	0
bornmapp68	Borner, K., Penumarthy, S., Meiss, M., and Ke, W.M., 2006. Mapping the diffusion of scholarly knowledge among major U.S. research institutions. *Scientometrics,* 68 (3), 415-426.	4	8
boyamapp101	Boyack, K.W., 2004. Mapping knowledge domains: characterizing PNAS. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 5192-5199.	1	18
boyaindi54	Boyack, K.W. and Borner, K., 2003. Indicator-assisted evaluation and funding of research: visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology,* 54 (5), 447-461.	2	18
boyamapp2007	Boyack, K.W., Borner, K., and Klavans, R., 2007. Mapping the structure and evolution of chemistry research. *11th International Conference of the International Society for Scientometrics and Informetrics,* 25-27 June 2007 Madrid. Leuven: ISSI, 45-60.	3	2
boyamapp79	Boyack, K.W., Borner, K., and Klavans, R., 2009. Mapping the structure and evolution of chemistry research. *Scientometrics,* 79 (1), 45-60.	3	3
boyameas2	Boyack, K.W. and Klavans, R., 2008. Measuring science-technology interaction using rare inventor-author names. *Journal of Informetrics,* 2 (3), 173-182.	2	0
boyamapp64	Boyack, K.W., Klavans, R., and Borner, K., 2005. Mapping the backbone of science. *Scientometrics,* 64 (3), 351-374.	3	73
boyaeval72	Boyack, K.W. and Rahal, N., 2005. Evaluation of laboratory directed research and development investment areas at Sandia. *Technological Forecasting and Social Change,* 72 (9), 1122-1136.	2	0
boyadoma53	Boyack, K.W., Wylie, B.N., and Davidson, G.S., 2002. Domain visualization using VxInsight (R) for science and technology management. *Journal of the American Society for Information Science and Technology,* 53 (9), 764-774.	3	43

boyainfo2539	Boyack, K.W., Wylie, B.N., and Davidson, G.S., 2002. Information visualization, human-computer interaction, and cognitive psychology: domain visualizations. *In:* K. Borner and C. Chen, eds. *Visual Interfaces to Digital Libraries (Lecture Notes in Computer Science v. 2539).* Berlin: Springer-Verlag, 145-158.	3	1
chendete2004	Chen, C.M., 2004. Detecting and mapping thematic changes in transient networks. *In:* E. Banissi, *et al*., eds. *8th International Conference on information Visualisation*,14-16 July 2004 London. Los Alamitos, CA: IEEE, 1023-1032.	1	0
chensear101	Chen, C.M., 2004. Searching for intellectual turning points: progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5303-5310.	1	48
chenmeas5669	Chen, C.M., 2005. Measuring the movement of a research paradigm. *In:* R.F. Erbacher, *et al*., eds. *Conference on Visualization and Data Analysis 2005 (Proceedings of the SPIE v. 5669)*, 17-18 January 2005 San Jose, CA. Bellingham, WA: SPIE, 63-76.	1	3
chencite57	Chen, C.M., 2006. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57 (3), 359-377.	1	68
chenholi25	Chen, C.M., 2007. Holistic sense-making: conflicting opinions, creative ideas, and collective intelligence. *Library Hi Tech*, 25 (3), 311-327.	1	0
chenfrom14	Chen, C.M. and Borner, K., 2005. From spatial proximity to semantic coherence: a quantitative approach to the study of group dynamics in collaborative virtual environments. *Presence - Teleoperators and Virtual Environments*, 14 (1), 81-103.	2	0
chenunde2005	Chen, C.M., Chen, Y.N., and Maulitz, R.C., 2005. Understanding the evolution of NSAID: a knowledge domain visualization approach to evidence-based medicine. 9th International Conference on Information Visualisation, 6-8 July 2005 London. Los Alamitos, CA: IEEE, 945-952.	3	0
chentrac59	Chen, C.M. and Hicks, D., 2004. Tracing knowledge diffusion. *Scientometrics*, 59 (2), 199-211.	2	12
chensema2003	Chen, C.M. and Lobo, N., 2003. Semantically modified diffusion limited aggregation for visualizing large-scale networks. *In:* F. Titsworth, *et al*., eds. 7th I*nternational Conference on Information Visualization*, 16-18 July 2003 London. Los Alamitos, CA: IEEE, 576-581.	2	0

chenmapp39	Chen, C.M., McCain, K., White, H., and Lin, X., 2002. Mapping Scientometrics (1981-2001). *In:* E.G. Toms, ed. *ASIST 2002: 65th ASIST Annual Meeting (v. 39),* 18-21 November 2002 Philadelphia. Medford, NJ: Information Today, 25-34.	4	1
chenvisu2003	Chen, C.M. and Morris, S., 2003. Visualizing evolving networks: minimum spanning trees versus pathfinder networks. *INFOVIS 2002: IEEE Symposium on Information Visualization*, 19-21 October 2002 Seattle. New York: IEEE, 67-74.	2	10
chenthem67	Chen, C.M., Song, I.Y., Yuan, X.J., and Zhang, J., 2008. The thematic and citation landscape of data and knowledge engineering (1985-2007). *Data & Knowledge Engineering,* 67 (2), 234-259.	4	1
chentren2007	Chen, C.M., Song, I.Y., and Zhu, W.Z., 2007. Trends in conceptual modeling: citation analysis of the ER conference papers (1979-2005). *11th International Conference of the International Society for Scientometrics and Informetrics,* 25-27 June 2007 Madrid. Leuven: ISSI, 189-200.	3	1
chenvisu2009	Chen, C.M., Zhang, J., and Vogeley, M S., 2009. Visual analysis of scientific discoveries and knowledge diffusion. *In:* B. Larsen and J. Leta, eds. *12th International Conference of the International Society for Scientometrics and Informetrics*, 14-17 July 2009 Rio de Janeiro. Leuven: ISSI, 874-885.	3	0
chendeli2007	Chen, C.M., Zhang, J., Zhu, W.Z., and Vogeley, M., 2007. Delineating the citation impact of scientific discoveries. *7th ACM/IEEE Joint Conference on Digital Libraries*, 18-23 June 2007 Vancouver. New York: ACM, 19-28.	4	0
garffrom48	Garfield, E., 1998. From citation indexes to informetrics: is the tail now wagging the dog? *Libri,* 48 (2), 67-80.	1	37
garfhist30	Garfield, E., 2004. Historiographic mapping of knowledge domains literature. *Journal of Information Science,* 30 (2), 119-145.	1	28
garffrom2007	Garfield, E., 2007. From the science of science to scientometric: visualizing the history of science with HistCite software. *11th International Conference of the International Society for Scientometrics and Informetrics,* 25-27 June 2007 Madrid. Leuven: ISSI, 21-26.	1	0
garffrom3	Garfield, E., 2009. From the science of science to scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics,* 3 (3), 173-179.	1	1
herryear2008	Herr, B.W., *et al.*, 2008. 113 years of physical review: Using flow maps to show temporal and topical citation patterns. *In:* E. Banissi, *et al.*, eds. *12th International Conference Information Visualisation*, 9-11 July 2008 London. Los Alamitos, CA: IEEE, 421-426.	5	0

hookdoma2007	Hook, P.A., 2007. Domain maps: purposes, history, parallels with cartography, and applications. *In:* E. Banissi, *et al*., eds. *11th international conference information visualization.* Los Alamitos, CA: IEEE, 442-446.	1	0
hookvisu2007	Hook, P.A., 2007. Visualizing the topic space of the united states supreme court. *11th International Conference of the International Society for Scientometrics and Informetrics,* 25-27 June 2007 Madrid. Leuven: ISSI, 387-396.	1	0
huanlong5	Huang, Z., *et al*., 2003. Longitudinal patent analysis for nanoscale science and engineering: country, institution and technology field. *Journal of Nanoparticle Research,* 5 (3-4), 333-363.	7	34
klaviden57	Klavans, R. and Boyack, K.W., 2006. Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology,* 57 (2), 251-263.	2	34
klavquan68	Klavans, R. and Boyack, K.W., 2006. Quantitative evaluation of large maps of science. *Scientometrics,* 68 (3), 475-499.	2	14
klavis2007	Klavans, R. and Boyack, K.W., 2007. Is there a convergent structure of science? A comparison of maps using the ISI and Scopus databases. *11th International Conference of the International Society for Scientometrics and Informetrics,* 25-27 June 2007 Madrid. Leuven: ISSI, 437-448.	2	3
kostscie48	Kostoff, R.N. and Scaller, R.R., 2001. Science and technology roadmaps. *IEEE Transactions on Engineering Management,* 48 (2), 132-143.	1	95
laroscho2007	LaRowe, G., *et al*., 2007. The scholarly database and its utility for scientometrics research. *11th International Conference of the International Society for Scientometrics and Informetrics,* 25-27 June 2007 Madrid. Leuven: ISSI, 457-462.	5	2
laroscho79	LaRowe, G., *et al*., 2009. The scholarly database and its utility for scientometrics research. *Scientometrics,* 79 (2), 219-234.	5	1
lilong10	Li, X., *et al*., 2008. A longitudinal analysis of nanotechnology literature: 1976-2004. *Journal of Nanoparticle Research,* 10, 3-22.	6	2
linanal3	Lin, J.M., *et al.,* 2008. An analysis of the abstracts presented at the annual meetings of the society for neuroscience from 2001 to 2006. *PloS One,* 3 (4), 2052.	6	2
linnew2007	Lin, X., 2007. New visual interfaces for author co-citation mapping. *11th International Conference of the International Society for Scientometrics and Informetrics*, 25-27 June 2007 Madrid. Leuven: ISSI, 882-883.	1	0
linreal39	Lin, X., White, H.D., and Buzydlowski, J., 2003. Real-time author co-citation mapping for online searching. *Information Processing & Management,* 39 (5), 689-706.	3	21

liuvisu62	Liu, Z., 2005. Visualizing the intellectual structure in urban studies: a journal co-citation analysis (1992-2002). *Scientometrics*, 62 (3), 385-402.	1	3
liumapp31	Liu, Z. and Wang, C.Z., 2005. Mapping interdisciplinarity in demography: a journal network analysis. *Journal of Information Science*, 31 (4), 308-316.	2	4
mahevisu2009	Mahesh, S., Trumbach, C.C., and Walsh, K.R., 2009. Visualizing technology mining results on life cycle axes: a text data mining study of publication trends in server virtualization from scientific research to business process improvement. *Conference of the Portland International Center for Management of Engineering and Technology*, 2-6 August 2009 Portland, OR. New York: IEEE, 2380-2393.	3	0
maricont52	Marion, L.S. and McCain, K.W., 2001. Contrasting views of software engineering journals: author cocitation choices and indexer vocabulary assignments. *Journal of the American Society for Information Science and Technology*, 52 (4), 297-308.	2	11
mccabiot46	McCain, K.W., 1995. Biotechnology in context - a database-filtering approach to identifying core and productive non-core journals supporting multidisciplinary research-and-development. *Journal of the American Society for Information Science*, 46 (4), 306-317.	1	23
mccaanal2007	McCain, K.W., 2007. Analysing influence over time: an historiographic mapping of the research of Conrad Hal Waddington (1905-1975).*11th International Conference of the International Society for Scientometrics and Informetrics*, 25-27 June 2007 Madrid. Leuven: ISSI, 558-567.	1	0
mccarela2007	McCain, K.W., 2007. The relationship between influence and image: two views of the oeuvre of Conrad Hal Waddington using historiographic mapping and author tri-citation image analysis. *11th International Conference of the International Society for Scientometrics and Informetrics*, 25-27 June 2007 Madrid. Leuven: ISSI, 890-891.	1	0
mcmimapp38	McMillan, G.S., 2008. Mapping the invisible colleges of R&D management. *R & D Management*, 38 (1), 69-83.	1	0
morrdiva43	Morris, S., *et al*., 2002. Diva: a visualization system for exploring document databases for technology forecasting. *Computers & Industrial Engineering*, 43 (4), 841-862.	5	17
morrsom2001	Morris, S.A., Wu, Z., and Yen, G., 2001. A SOM mapping technique for visualizing documents in a database. *International Joint Conference on Neural Networks*, 15-19 July 2001 Washington, DC. New York: IEEE, 1914-1919.	3	0

morrcros101	Morris, S.A. and Yen, G.G., 2004. Crossmaps: visualization of overlapping relationships in collections of journal papers. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 5291-5296.	2	11
morrvisu2000	Morris, T., 2000. Visualizing the structure of medical informatics using term co-occurrence analysis. *In:* D.H. Kraft, ed. *Proceedings of the 63rd ASIS Annual Meeting,* 12-16 November Chicago: Medford, NJ: Information Today, 209-218.	1	0
morrvisu38	Morris, T., 2001. Visualizing the structure of medical informatics using term co-occurrence analysis: II. INSPEC perspective. *In:* E. Aversa and C. Manley, eds. *Proceedings of the 64th ASIST Annual Meeting (v. 38),* 3-8 November 2001 Washington, DC. Medford: Information Today, 489-497.	1	0
neruinte29	Nerur, S.P., Rasheed, A.A., and Natarajan, V., 2008. The intellectual structure of the strategic management field: an author co-citation analysis. *Strategic Management Journal,* 29 (3), 319-336.	3	5
noelvisu2002	Noel, S., Chu, C.H., and Raghavan, V., 2002. Visualization of document co-citation counts. *Sixth International Conference on Information Visualisation,* 10-12 July 2002 London. Los Alamitos, CA: IEEE, 691-696.	3	0
noelvisu2001	Noel, S., Raghavan, V., and Chu, C.H.H., 2001. Visualizing association mining results through hierarchical clusters. *IEEE International Conference on Data Mining*, 29 November-2 December 2001 San Jose, CA. Los Alamitos, CA: IEEE, 425-432.	3	0
pajwmapp2005	Pajwani, G. and Mccain, K.W., 2005. Mapping the "human" side of computing: an author co-citation analysis of the interrelationships between ergonomics, human-computer interaction and human factors research. *10th International Conference of the International Society for Scientometrics and Informetrics,* 24-28 July 2005 Stockholm. Stockholm: Karolinska University Press, 683-684.	2	0
pereinte2008	Perer, A. and Shneiderman, B., 2008. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. *26th Annual Chi Conference on Human Factors in Computing Systems*, 5-10 April 2008 Florence. New York: ACM, 265-274.	2	1
perrscho49	Perry, C.A. and Rice, R.E., 1998. Scholarly communication in developmental dyslexia: influence of network structure on change in a hybrid problem area. *Journal of the American Society for Information Science,* 49 (2), 151-168.	2	16

porthow11	Porter, A. L. and Youtie, J., 2009. How interdisciplinary is nanotechnology? *Journal of Nanoparticle Research*, 11 (5), 1023-1041.	2	1
reidmapp3495	Reid, E. and Chen, H.C., 2005. Mapping the contemporary terrorism research domain: researchers, publications, and institutions analysis. *In:* P. Kantor, *et al*., eds. *Intelligence and Security Informatics (Lecture Notes in Computer Science v. 3495),* 19-20 May 2005 Atlanta, GA. Berlin: Springer-Verlag, 322-339.	2	3
reidmapp65	Reid, E.F. and Chen, H.C. 2007. Mapping the contemporary terrorism research domain. *International Journal of Human-Computer Studies,* 65 (1), 42-56.	2	7
reveinfo107	Revere, D., Fuller, S., Bugni, P.F., and Martin, G.M., 2004. An information extraction and representation system for rapid review of the biomedical literature. *In:* M. Fieschi, E. Coiera, and Y.C.J. LI, eds. *11th World Congress on Medical Informatics (Studies in Health Technology and Informatics v. 107)*, 7-11 September 2004 San Francisco. Amsterdam: IOS Press, 788-792.	4	0
skupworl101	Skupin, A., 2004. The world of geography: visualizing a knowledge domain with cartographic means. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 5274-5278.	1	13
skupdisc3	Skupin, A., 2009. Discrete and continuous conceptualizations of science: implications for knowledge domain visualization. *Journal of Informetrics,* 3 (3), 233-245.	1	2
smalupda38	Small, H., 1997. Update on science mapping: creating large document spaces. *Scientometrics,* 38 (2), 275-293.	1	36
smalvisu50	Small, H., 1999. Visualizing science by citation mapping. *Journal of the American Society for Information Science,* 50 (9), 799-813.	1	139
smaltrac2005	Small, H., 2005. Tracking and predicting growth areas in science. *10th International Conference of the International Society for Scientometrics and Informetrics*, 24-28 July 2005 Stockholm. Stockholm: Karolinska University Press, 13-23.	1	0
smaltrac68	Small, H., 2006. Tracking and predicting growth areas in science. *Scientometrics,* 68 (3), 595-610.	1	22
smalmaps2009	Small, H., 2009. Maps of science as interdisciplinary discourse: co-citing contexts and the role of analogy. *In:* B. Larsen and J. Leta, eds. *12th International Conference of the International Society for Scientometrics and Informetrics,* 14-17 July 2009 Rio de Janeiro. Leuven: ISSI, 270-274.	1	0

smalcita2007	Small, H. and Upham, P., 2007. Citation structure of an emerging research area: organic thin film transistors. *11th International Conference of the International Society for Scientometrics and Informetrics,* 25-27 June 2007 Madrid. Leuven: ISSI, 718-725.	2	0
songsema38	Song, M. and Galardi, P., 2001. Semantic relationships between highly cited articles and citing articles in information retrieval. *In:* E. Aversa and C. Manley, eds. *Proceedings of the 64th ASIST Annual Meeting (v. 38)*, 3-8 November 2001 Washington, DC. Medford, NJ: Information Today, 171-181.	2	0
tangevol41	Tang, R., 2004. Evolution of the interdisciplinary characteristics of information and library science. *In:* J.B. Bryans ed. *Proceedings of the 67th ASIS&T Annual Meeting (v. 41), 12-17 November 2004* Providence, RI. Medford, NJ: Information Today, 54-63.	1	1
thakcont5295	Thakur, S., *et al.*, 2004. Content coverage of animal behavior data. *In:* R.F. Erbacher, *et al*., eds. *Visualization and Data Analysis,* 19-20 January 2004 San Jose, CA. Bellingham, WA: SPIE, 305-311.	5	0
tsaiinfo2001	Tsai, B.S., 2001. Information mapping, charting and visualization for total quality knowledge management: constructing and assessing a web-coordinated experts map-chart/depository/query-report system through citation data mining and information landscaping. *8th International Conference on Scientometrics and Informetrics*, 16-20 July 2001 Sydney. Kensington: University of New South Wales, 691-704.	1	0
tsaiinfo39	Tsai, B.S., 2003. Information landscaping: information mapping, charting, querying and reporting techniques for total quality knowledge management. *Information Processing & Management*, 39 (4), 639-664.	1	1
whitend38	White, H., Lin, X., and Buzydlowski, J., 2001. The endless gallery: visualizing authors' citation images in the humanities. *In:* E Aversa and C. Manley, eds. *Proceedings of the 64th ASIST Annual Meeting (v. 38),* 3-8 November 2001 Washington, DC. Medford, NJ: Information Today, 182-189.	3	0
whitauth51	White, H.D., 2001. Author-centered bibliometrics through CAMEOs: characterizations automatically made and edited online. *Scientometrics*, 51 (3), 607-637.	1	15
whitpath54	White, H.D. 2003. Pathfinder networks and author cocitation analysis: a remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology,* 54 (5), 423-434.	1	65

whitcomb58a	White, H.D., 2007. Combining bibliometrics, information retrieval, and relevance theory, part 1: first examples of a synthesis. *Journal of the American Society for Information Science and Technology*, 58 (4), 536-559.	1	3
whitcomb58b	White, H.D., 2007. Combining bibliometrics, information retrieval, and relevance theory, part 2: some implications for information science. *Journal of the American Society for Information Science and Technology*, 58 (4), 583-605.	1	3
whitco2000	White, H.D., Buzydlowski, J., and Lin, X., 2000. Co-cited author maps as interfaces to digital libraries: designing pathfinder networks in the humanities. *In:* E. Banissi, *et al*., eds. *IEEE International Conference on Information Visualisation*, 19-21 July 2000 London. Los Alamitos, CA: IEEE, 25-30.	3	4
whituser101	White, H.D., Lin, X., Buzydlowski, J.W., and Chen, C.M., 2004. User-controlled mapping of significant literatures. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5297-5302.	4	11
whitvisu49	White, H.D. and McCain, K.W., 1998. Visualizing a discipline: an author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49 (4), 327-355.	2	235
yenself2006	Yen, G.G. and Wu, Z., 2006. A self-organizing map based approach for document clustering and visualization. *IEEE International Joint Conference on Neural Network Proceedings*, 16-21 July 2006 Vancouver. New York: IEEE, 3279-3286.	2	0
zhanvisu15	Zhang, J., Chen, C.M. and Li, J.X. 2009. Visualizing the intellectual structure with paper-reference matrices. *IEEE Transactions on Visualization and Computer Graphics*, 15 (6), 1153-1160.	3	0

Appendix C

SAS Program for Normality Testing

The SAS program below was utilized to perform several normality tests, including the Shapiro-Wilk test, to determine normality of the average distance between co-authors within years. The program was derived from documentation and examples from SAS OnlineDoc 9.1.3 (SAS Institute 2008). The p values for the Shapiro-Wilk test are summarized in Table 8.

```
options ls=100 ps=56 nonumber nodate;

data one;
title1 'Distance - Normality Test';
input year distance;
cards;
2001  1024
2001  13
2001  0
2001  0
2001  0
2001  0
2002  1024
2002  0
2002  0
2002  0
2002  0
2003  869
2003  1134
2003  0
2003  1213
2003  325
2003  0
2004  0
2004  662
2004  0
2004  0
2004  0
2004  0
2005  0
2005  0
2005  1430
2005  0
2005  603
2005  6
2005  1433
2005  0
2005  0
2006  0
2006  0
2006  1725
2006  1725
2006  0
```

```
2007  1430
2007  0
2007  0
2007  1725
2007  0
2007  1843
2007  0
2008  1718
2008  67
2008  0
2008  388
2008  496
2008  129
2008  0
2009  210
2009  281
2009  0
2009  0
2009  0
2009  1430
2009  0
2009  0
2009  0
2009  0
2009  0
;
run;

proc print;
Run;

proc sort; by year;
run;

proc univariate normal; by year;
var distance;
run;
```

Appendix D

SAS Program for Kruskal-Wallis Test

The SAS program below performs the Kruskal-Wallis Test to determine differences among means. The program was derived from documentation and examples from SAS OnlineDoc 9.1.3 (SAS Institute 2008). The SAS output is presented below and discussed in the Results section.

```
options ls=100 ps=56 nonumber nodate;
data one;
title1 'Distance - Kruskal-Wallis Test';
input year distance;
cards;
2001    1024
2001    13
2001    0
2001    0
2001    0
2001    0
2002    1024
2002    0
2002    0
2002    0
2002    0
2003    869
2003    1134
2003    0
2003    1213
2003    325
2003    0
2004    0
2004    662
2004    0
2004    0
2004    0
2004    0
2005    0
2005    0
2005    1430
2005    0
2005    603
2005    6
2005    1433
2005    0
```

```
2005    0
2006    0
2006    0
2006    1725
2006    1725
2006    0
2007    1430
2007    0
2007    0
2007    1725
2007    0
2007    1843
2007    0
2008    1718
2008    67
2008    0
2008    388
2008    496
2008    129
2008    0
2009    210
2009    281
2009    0
2009    0
2009    0
2009    1430
2009    0
2009    0
2009    0
2009    0
2009    0
;
run;

proc print;
run;

proc sort; by year;
run;

proc npar1way wilcoxon data=one;
 class year;
 var distance;
run;
```

Distance - Kruskal-Wallis Test

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable distance
Classified by Variable year

year	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
2001	6	165.50	189.00	37.267481	27.583333
2002	5	126.50	157.50	34.322810	25.300000
2003	6	236.00	189.00	37.267481	39.333333
2004	6	143.00	189.00	37.267481	23.833333
2005	9	292.00	283.50	44.403744	32.444444
2006	5	177.00	157.50	34.322810	35.400000
2007	7	253.00	220.50	39.892455	36.142857
2008	7	268.00	220.50	39.892455	38.285714
2009	11	292.00	346.50	48.155049	26.545455

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	7.1500
DF	8
Pr > Chi-Square	0.5205

Conclusion: Do not reject Ho. (Ho: The annual mean distances are the same.)

References

Adair, W.C., 1955. Citation indexes for scientific literature. *American Documentation*, 6, 31-32.

Adams, J.D., Black, G.C., Clemmons, J.R., and Stephan, P.E., 2005. Scientific teams and institutional collaborations: evidence from U.S. universities, 1981-1999. *Research Policy*, 34 (3), 259-285.

Allen, G., 1960. Citation network. *In:* E. Garfield, ed. *Citation Index to Genetics and General Science Literature*. Research Proposal by the Institute for Scientific Information. Submitted July 15, 1960 to the National Science Foundation.

Batty, M., 2003. The geography of scientific citation. *Environment and Planning A*, 35 (5), 761-765.

Bertin, J., 1981. *Graphics and graphic information processing*. Berlin: Walter de Gruyter.

Bertin, J., 1983. *Semiology of graphics: diagrams, networks, maps*. Madison, WI: University of Wisconsin.

Borner, K., 2010. *Atlas of science: visualizing what we know*. Cambridge, MA: MIT Press.

Borner, K., Penumarthy, S., Meiss, M., and Ke, W., 2006. Mapping the diffusion of scholarly knowledge among major U.S. research institutions. *Scientometrics*, 68 (3), 415-426.

Cairncross, F., 2001. *The death of distance: how the communications revolution is changing our lives*. Boston: Harvard Business School Press.

Carvalho R. and Batty, M., 2006. The geography of scientific productivity: scaling in US computer science. *Journal of Statistical Mechanics - Theory and Experiment*, P10012.

CD Light, 2010. *Free ZIP code lookup*. Available from: http://zipinfo.com/search/zipcode.htm [Accessed 16 January 2011].

Chen, C., 2007. Holistic sense-making: conflicting opinions, creative ideas, and collective intelligence. *Library Hi Tech*, 25 (3), 311-327.

Chen, C.M., Zhang, J., Zhu, W., and Vogeley, M., 2007. Delineating the citation impact of scientific discoveries. *7th ACM/IEEE Joint Conference on Digital Libraries*, 18-23 June 2007 Vancouver. New York: ACM, 19-28.

Chen, C.M., Song, I.Y., Yuan, X., and Zhang, J., 2008. The thematic and citation landscape of data and knowledge engineering (1985-2007). *Data & Knowledge Engineering*, 67 (2), 234-259.

Dodge, M., McDerby, M., and Turner, M., eds., 2008. *Geographic visualization: concepts, tools and applications*. Chichester, England: John Wiley & Sons.

Dykes, J., 2008. Geovisualization. *In*: K.K. Kemp, ed. *Encyclopedia of geographic information science*. Thousand Oaks, CA: Sage, 209-213.

ESRI, 2004. Create polyline shapefile from table sample. Available from: http://edn.esri.com/index.cfm?fa=codeExch.sampleDetail&pg=/arcobjects/9.1/Samples/ArcCatalog/Commands_and_Tools/Make_Line_Feature_Class/MakeLineFeatureClass.htm [Accessed 22 January 2011].

ESRI [CD-ROM], 2008. *ESRI Data & Maps Series*. Redlands, CA: ESRI.

Garfield, E., 1955. Citation indexes for science. *Science*, 122 (3159) 108-111.

Gastner, M.T. and Newman, M.E., 2004. Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences*, 101 (20), 7499-7504.

Glanzel, W., 2001. National characteristics in international scientific co-authorship relations. *Scientometrics*, 51 (1), 69-115.

Griffin, A.L., MacEachren, A.M., Hardisty, F., Steiner, E., and Li, B., 2006. A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Annals of the Association of American Geographers*, 96 (4), 740-753.

Hoekman, J., Frenken, K., and Tijssen, R.J.W., 2010. Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39 (5), 622-673.

LaRowe, G., Ambre, S., Burgoon, J., Ke, W., and Borner, K., 2009. The scholarly database and its utility for scientometrics research. *Scientometrics*, 79 (2), 219-234.

Leydesdorff, L. and Persson, O., 2010. Mapping the geography of science: distribution patterns and networks of relations among cities and institutes. *Journal of the American Society for Information Science and Technology*, 61 (8), 1622-1634.

Lin, J.M., Bohland, J.W., Andrews, P., Burns, G.A., Allen, C.B., and Mitra, P.P., 2008. An analysis of the abstracts presented at the annual meetings of the society for neuroscience from 2001 to 2006. *PloS One*, 3 (4), 2052.

MacEachren, A.M., 2004. *How maps work: representation, visualization, and design*. NY: Guilford.

MacEachren**,** A.M. and Kraak, M.J., 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science*, 28 (1), 3-12.

Maggioni, M.A. and Uberti, T.E., 2009. Knowledge networks across Europe: which distance matters? *Annals of Regional Science*, 43 (3), 691-720.

Matthiessen, C.W. and Schwartz, A.W., 1999. Scientific centres in Europe: an analysis of research strength and patterns of specialisation based on bibliometric indicators. *Urban Studies*, 36 (3), 453-477.

Meiss, M., 2005. *Chizu tutorial*. Available from: http://iv.slis.indiana.edu/lm/lm-chizu.html. [Accessed 10 June 2010].

Mothe, J., Chrisment, C., Dkaki, T., Dousset, B., and Karouach, S., 2006. Combining mining and visualization tools to discover the geographic structure of a domain. *Computers, Environment, and Urban Systems*, 30 (4), 460-484.

Old, L.J., 2001. Utilizing spatial information systems for non-spatial-data analysis. *Scientometrics*, 51 (3), 563-571.

SAS Institute, 2008. *SAS OnlineDoc 9.1.3*. Available from: http://support.sas.com/onlinedoc/913/docMainpage.jsp [Accessed 10 June 2011].

Skupin, A., 2002. A cartographic approach to visualizing conference abstracts. *IEEE Computer Graphics and Applications,* 22 (1), 50-58.

Skupin, A., 2004. The world of geography: visualizing a knowledge domain with cartographic means. *Proceedings of the National Academy of Sciences*, 101 (suppl 1), 5274-5278.

Slocum, T.A., McMaster, R.B., Kessler, F.C., and Howard, H.H., 2005. *Thematic cartography and geographic visualization* (2nd ed.). Upper Saddle River, NJ: Pearson.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography,* 46 (2), 234-240.

Thomson Reuters., 2010. *Web of Science* [online]. Philadelphia, Thomson Reuters. Available from: apps.webofknowledge.com [Accessed 7 November 2010].

Tufte, E., 1990. *Envisioning information*. Cheshire, CT: Graphic Press.

Tversky, B., Morrison, J.B., and Betrancourt, M., 2002. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57 (4), 247-262.

Waltman, L., Tijssen, R.J.W., and Van Eck, N.J., 2011. Globalisation of science in kilometres. *Journal of Informetrics*, 5 (4), 574-582.